

Db	Sequence	Accession	Source
575	ATGCTGACCTTCAGACACTACTGGTGCACCTTCAGGCTGCATATACCCAGGCACTGCGAA	634	
543	GAGCACATCATTTCTGAAACAGCAACTTGTGACACTGTGCGCTCAGCAATTCATTAACCTTTGG	602	
635	GCCTATCTCTCTCTGAGCAACAGACTTGTGACTGTGGTTTGTGATTCACAAATTCGG	694	
603	ATGCAATGGAGGTTTGGCTTCCAGGCTTTCGAATACCTTAAGTACATGAGGAGCATCGA	662	
695	ATGCAACGGAGGCTTCCATCCAGGCTTTCGAATACATCAAAATACAAATGGTGGCTTTGA	754	
663	CACAGAACAGACTTATCCATACCTTGGTGTCAATGGTATCTGCAACTTCAAGCAGAGAA	722	
755	CACAGAGAACTTACCTTACCAAGGTGTCAATGGAATCTCCAAATTCAGAAATGAGAA	814	
723	TGTTGGTGTCAAGCTCATTCATTCGATTAACATACCTGGTGGTCTGAGGATGAGTTGAA	782	
815	TGTTGGAGTCAAGGTTTGGACTCGGTTAAACATACCTGGTGGTCTGAGGATGAGTTGAA	874	
783	GCATGCAGTGGGCTTGGTGGTCCAGTTAGCGTTGCATTTGAGGTTGTGAAAGGTTTCAA	842	
875	GGATGCTGTTGGTCTGGTGGTCCAGTTAGTGTTCCTTCGAGGTGATCACTGGTTTCAG	934	
843	TCTGTACAGAAAGGTGTATACAGCAGTGCACACCTGTGGAAGAGATCCAAATGGATGTGAA	902	
935	GCTGTACAGAGCGAGTTTACATAGCAGCACTTGTGGAACACTACCGATGGATGTGAA	994	
903	CCAGCAGTTCTTCCCGTGGTGTATGAGAGTGCAGAGCGGATTCCTTTATGGCTCATCAA	962	
995	CCAGCTGTTCTGGCTGTTGGCTACCGTGTGGAAGATGGTGTACCTACTGGCTCATCAA	1054	
963	GAACTCATGGGTCACAAATGGGGTGACAAATGGCTACTTTAAGATGGAATCGGCAAGAA	1022	
1055	GAACTCATGGGCGCTGACTGGGTTGATGAGGTTTACTTCAAGATGGAAATGGGCAAGAA	1114	
1023	CATGTGTGGTGTGCAACTTGGCGCATCTTATCCCAATTCGTGGC	1064	
1115	CATGTGGGCTGTTGCTACGTGTGCATCTTACCTTATGTCG	1156	

RESULT 4

MZBCYPA

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

MZBCYPA

Corn mRNA for cysteine proteinase, clone CCP2, complete cds.

D45403

GI:1688044

cysteine proteinase.

Zea mays

Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 1382)

Domoto, C., Watanabe, H., Abe, M., Abe, K. and Arai, S.

Isolation and characterization of two distinct cDNA clones encoding corn seed cysteine proteinases

Biochim. Biophys. Acta 1263 (3), 241-244 (1995)

96004895

7548211

2 (bases 1 to 1382)

Domoto, C.

Direct Submission

Submitted (31-JAN-1995) Chieko Domoto, Atom Junior College; Ohtsuka 1-5-2, Bunkyo-ku, Tokyo 112, Japan

(Tel:03-3941-8161(ex.575), Fax:03-3945-1836)

On Nov 27, 1996 this sequence version replaced gi:644489.

Sequence updated (26-Oct-1996) by: Chieko Domoto.

Location/Qualifiers

1..1382

/organism="Zea mays"

/mol_type="mRNA"

/db_xref="taxon:4577"

/clone="CCP2"

[illegible]

191 GGAAGAGTACGGATCGAGAGGAGATCAAGAGAGGTTCCGGATCTTCGTGAGAAATC 250
232 GCAAGAGCTACGGAGCGCGGAGTGAAGAGCGGTTCCGGATCTTCGTGAGAGCC 291
251 TAGCGTTATCCGCTCCACATTAATCGAAGAGATCTGCTGATACCTAGGAATCAACCAAT 310
292 TCGAGAGAGTCCGCTCCACCAACCGAAGAGGCTCTCTCAAGCTCGGATCAACCGAT 351
311 TCGCGACCTGACCTCGGAGGAATTCGGACCAATCGCCTTGTGCGCGCAGAACTGCT 370
352 TCTCTGACATGACCTGGAGGAGTTCCAGGCCACCAAGCTTGGCGCGCAGACCTGCT 411
371 CGGCGACTGCGCATGGAACCAACCGGTT---TGTGATGCGGTCTTCCTGTAAACGAGG 427
412 CGGCCACGCTCCGCGGCAACCACTGATGCGGAGCGCAACGCGCTCCGCGAGCAAAAG 471
428 ATTGAGGAGCAAGGATAGTAGGCTGTAAAGGACCAAGGAAGCTGTGGATCTTGTCT 487
472 ACTGGAGGAGACTGGGATCGTTAGCCCTGTGAAGACCAAGGCAAGCTGCGGCTCTGCT 531
488 GGACTTTCACTACTGAGGAGCTAGAGGCTGCATATACACAGCTAACTGGAAGAGCA 547
532 GGACCTTCAGCACTACTGAGCACTTGAAGGAGCATATACTCAGGCCCACTGGAAGAAC 591
548 CATCAATATCGAAGCACTTGTGACTGTGCTCAGCAATCAATAAATTTGGATGCA 607
592 TCTCTCTTCTGAGCAACAGCTGGTGTGCTGCGGTGCATACAACTTTGGATGCA 651
608 ATGAGGTTTCCCTTCCCAAGCCTTTGAATACGTTAAGTAACTAGGAGGATCGACACAG 667
652 ACGAGGCTTCCATCCCAAGCACTTGAATACAGTCAATGAGTGGCATTCAGACCG 711
668 AACAGCTATCCATACCTTGGTGTCAATGCTATCGCACTTCAAGCAGGAGATCTG 727
712 AGGAGCTTACCTTTACAAGGCTGTCAATGGGCTGTGCAATACAGGCTTGAATGCTG 771
728 GTGTCAAGGCTCAATGATTCGATAAACAATCAACCTCGGTGTGAGGATGAGTTGAAGCATG 787
772 CAGTTACAGTTTGGGACTCGGTCAACATCACTCTGAAATGCTGAGGATGAATGAAGATG 831
788 CAGTGGCTTGGTGGCTCAGTTAGCTGCTGATTTAGGTTGTGAAGGTTTCAATCTGT 847
832 CTGTGGGTTGGTTGGCCCGCTCAGTGTGCTTTGAGGTGATCGAGGTTTCAAGCAGT 891
848 ACAAGAAAGGTGTATACAGCACTGACCTGTGAAGAGATCCCAATGATGTGAACCAAG 907
892 ACAGAGCGGAGTTTACACTAGTGACCAATTTGTGGAATCTACTCTGACGATGTCAACCAAG 951
908 CAGTTCTTGGCTCGGTTATGAGTTCGAGGACGGGATTCCTTATTTGGCTCATCAAGAACT 967
952 CTGTTCTAGCTGTGGCTACGCTGTGAGAAATGGCGTCCCGTATTTGGCTCATCAAGAACT 1011
968 CATGGGTACAAATTTGGGCTGACATGCTACTTTAAGATGGAATCGGCAAGAACATGT 1027
1012 CATGGGCGCGACTGGGGTGAAGAGGTTTACTTCAAGATGGAATGGCAAGAACATGT 1071
1028 GTGTGTTTGCACCTTGGCACTTATCCCATTTGGCTG 1066
1072 GTGTGCTGCTACCTGGCGCTCTACCTATCTTTGGCGG 1110

RESULT 7
AR231148
LOCUS AR231148 1407 bp DNA linear PAT 20-DEC-2002
DEFINITION Sequence 107 from patent US 6451604.
ACCESSION AR231148
VERSION AR231148.1 GI:27271935
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1407)

AUTHORS Flinn, B. and Lasham, A.
TITLE Compositions affecting programmed cell death and their use in the modification of forestry plant development
JOURNAL Patent: US 6451604-A 107 17-SEP-2002;
FEATURES Location/Qualifiers
source 1. 1407
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 50.6%; Score 541.4; DB 6; Length 1407;
Best Local Similarity 70.5%; Pred. No. 4.2e-138;
Matches 738; Conservative 0; Mismatches 306; Indels 3; Gaps 1;
21 CTTCTACTCTTCTCTCGCCCTCTCCGTCGCGCTCGCGGTTTCGAAGAAGACAA 80
Db CGTCTCTCTCTCTCGCGCTCTCGTCTCCGCCCGCGGTCGAGCTTCGAGAGTCCAA 146
21 TCCATATCCCGTCCGTTACACA---AAGGCTTACTCGATTGAGCTTCGCGCTCTCGGCT 137
Db CCCCATCCGCTCTTCCCGACGCGGCTCCGCGACCTCGAGTCTCTCCATCGTCCAGAT 206
138 CTTTGGCAGTTGCGCGCAGCGCTTCACTTCCGACGTTTCCGCCGCGAGTACGGAGAG 197
207 CGTGGCGCGCACCCGCGCACGCTTCTCTTCCGCCGCTTCGCCAACAGGATGCGGAGAG 266
198 CTACGGATCGGAGGAGGATCAAGAAGAGGTTCCGGATCTTCTGTTGGAAATCTAGCGTT 257
267 GTACGAGACCGCGGAGGAGATCAAGCTCGGTTTCAGATCTTCAGGAGAGATCTCAAGT 326
258 TATCCGCTCCATATTCGAAGAGATCTGTCTGATACCTAGGAATCAACCAATTCGCCGA 317
327 GATCCGATCCACCAACAAGAGGCTTGCCTTACACCTTCGCTCGGTGTCAATTAAGTTTGTCA 386
318 CCGTACCTCGGAGGATTCGCGACCAATCGCTTGTGGCGCGCAGAACTGTCGCGCAC 377
387 TTGGAGCTGGAGAGGATTCAGGAGGCAAGATCGGAGTGTCAAAATGCTCTGCGCAC 446
378 TGGCATGGAACCAACCGGTTTGTGATGCGCTGCTTCTGTAACGAGGATGGAAGGA 437
447 CACCAAGGCAACCAACCAAGCTCACCGCAAGAGCTTCTCCCGAGATGAAAGACTGGAGAGA 506
438 GCAAGGATAGTGAAGCGCTTGAAGACCAAGAGAGTGTGATCTTGTGAGCTTTTCAG 497
507 AAGGCGATTTGAAGCGCAATTAAGATCAGGCGCACTGTGGATCTTGTGAGCTTTCAAG 566
498 TACTACTGAGCACTAGAGGCTGCATATACAGAGTAACTGGAAGAGGACATCATATATC 557
567 TACCCTGAGCTTTGAGGCTGCTTATCACAAGCAATTCGGGAAACAAATCTCTCTGTC 626
558 TGAACAGCAACTTGTGAGCTGCTCAGCATTAATTAATCTTGGATGCAATGGAGTTT 617
627 TGAAGCAGCACTGCTGAGCTGTGCTGGGCTTTTCAAACTTTGGATGTAGTGGTGGACT 686
618 GCCTTCCCAAGCTTTGAATACGTTAAGTACAAATGAGGATCGACACAGAAACAGACTTA 677
687 GCATCCCAAGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 746
678 TCCATACCTTGTGTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 737
747 TCCTTATACCGCAGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 806
738 CATTGATTCGATTAACATCACCTGCGTGTGAGGATGATGATGATGATGATGATGATGATGAT 797
807 GCTCGAGCTGTGTTAAATCACTTGGGCTGCTGAGGATGAACTTAAGCATGCAATGCTTCCCT 866
798 GGTGCGTCCAGTTAGCGTTGCAATTTGAGGTTTGTGAAGGTTTCAATCTGTACAAAGAGG 857
867 CGTCCGCGCAGTGAGTGTGCAATTCAGGTTCCAGGTTGTTGAAGAGCTTCAGATTTGACAAAGTCGGG 926
858 TGTATACAGAGTGCACCTGTGGAAGAGATCCAAATGATGATGATGATGATGATGATGATGATGAT 917
927 TGTCTACAGAGCGATACATGCGGTAGCACTTCCATGATGATGATGATGATGATGATGATGATGAT 986

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: March 29, 2004, 17:17:07 ; Search time 2963 Seconds
(without alignments)
15866.686 Million cell updates/sec

Title: US-10-087-714-1
Perfect score: 1071
Sequence: 1 atggcagctaagctctctt.....atcccatgtggtgtgtgag 1071

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0
Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pin.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sv.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Length	DB ID	Description
1	576.4	53.8	1486	8 AK066748 Oryza sat
2	569.6	53.2	1412	8 ZMSBE1 X99936 Z.mays mRNA
3	569.6	53.2	1442	6 A43549 Sequence 17
4	568	53.0	1382	8 MZECYPA A43549 Sequence 17
5	558.8	52.2	1468	8 RICOZC D90408 Oryza sativ
6	551.8	51.5	1345	8 LMU249847 Lottium mu
7	541.4	50.6	1407	6 AR231148 Sequence
8	518.8	48.4	1332	8 AF454960 Brassica
9	517.2	48.3	1444	8 PAU93166 Arachis
10	511.2	47.7	1340	8 AB032168 Nicotiana
11	488.4	45.6	1305	8 AF233883 Arabidops
12	488.4	45.6	1352	8 AY088662 Arabidops
13	488.4	45.6	1355	8 AF083703 Arabidops
14	486.8	45.5	1331	8 BT000676 Arabidops
15	486.8	45.5	1370	8 AF360273 Arabidops
16	486.2	45.4	1347	8 PHU31094 Arabidops
17	485.4	45.3	1388	6 A43538 Arabidops
18	485.4	45.3	1366	8 LECYP3GN Arabidops
19	485.2	45.3	1366	8 BT000674 Arabidops
20	485.2	45.3	1409	8 BT000673 Arabidops
21	474	44.3	1381	8 PSRNACP Arabidops
22	473.6	44.2	1484	8 CACG2 Arabidops
23	472.4	44.1	1361	8 PSA278699 Arabidops
24	467.6	43.7	1402	8 AY091771 Arabidops
25	344	32.1	937	8 AY383240 Glycine m
26	304	28.4	536	8 AF213939 Prunus du
27	280.2	26.2	677	6 AR231152 Sequence
28	259.8	25.2	1362	10 RNCATH Y00708 Rat mRNA fo
29	267	24.9	1360	10 RNCATH M38135 Rat catheps
30	262.6	24.5	1372	10 BC006878 Mus muscu
31	261.6	24.4	1149	4 SSAF001169 AF001169 Sus scrof
32	261	24.4	1381	10 MMCTSH U06119 Mus musculu
33	247.2	23.1	972	9 AF426248 Homo sapi
34	247.2	23.1	1008	9 AF426247 Homo sapi
35	247.2	23.1	1399	9 HSCATH X16832 Human mRNA
36	245.6	22.9	1470	9 BC002479 Homo sapi
37	244	22.8	1438	9 AK130158 Homo sapi
38	238.4	22.3	1106	6 AX774741 Sequence
39	238.4	22.3	1106	9 HSCATH X07549 Human mRNA
40	205.2	19.2	475	6 AR231161 Sequence
41	200.8	18.7	841	8 AF242733 Capsicum
42	197.6	18.5	535	8 NTPPCYSPT AF242733 Capsicum
43	182.8	17.1	498	10 MMU18464 Y11003 N.tabacum m
44	181.6	17.0	1168	3 AY336797 Arabidops
45	181.2	16.9	498	9 HOS18461 Y18461 Homo sapien

ALIGNMENTS

RESULT 1
AK066748
LOCUS
DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:J013074D19, full
linear 1486 bp mRNA
ACCESSION AK066748
VERSION AK066748.1 GI:32976766
KEYWORDS FLI_CDNA; CAP trapper.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1
AUTHORS The Rice Full-length cDNA Consortium, National Institute of

QY 918 CGTCGGTTATCGAGTCGAGGACGGATTCCTTATTGGCTCATCAAGAACTCATGGGTAC 977
 Db 987 TGTGGTTATGAGTGAAGATGGTGTTCCTTCTGGCTCATCAAGAAATTCCTGGGAGC 1046
 QY 978 AAATTGGGGTGACAAATGCTACTTTAAGATGGAACCTCGGCAAGAACATGTGTGTGTC 1037
 Db 1047 AGACTGGGGTGACCAAGGATACCTTCAGATGGAATGGAAGAACATGTGTGGATGCG 1106
 QY 1038 AACTTGCGCATCTTATCCCATGTGGC 1064
 Db 1107 TACTTGTGCATACACCTGTTGTGGC 1133

RESULT 8
 AF454960 1332 bp mRNA linear PLN 14-JAN-2002
 LOCUS
 DEFINITION Brassica oleracea senescence-associated cysteine protease (CPs)
 mRNA, complete cds.

ACCESSION AF454960
 VERSION AF454960.1 GI:18141288

KEYWORDS
 SOURCE
 ORGANISM

Brassica oleracea
 Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE
 1 (bases 1 to 1332)
 Coupe, S.A., Sinclair, B.K., Watson, L.M. and Eason, J.R.
 Cysteine proteases and broccoli senescence: cloning,
 characterization and contribution to the process
 Unpublished

JOURNAL
 REFERENCE
 2 (bases 1 to 1332)
 Coupe, S.A., Sinclair, B.K., Watson, L.M. and Eason, J.R.
 Direct Submission
 Submitted (05-DEC-2001) Crop and Food Research, Private Bag 11 600,
 Palmerston North 5301, New Zealand

FEATURES
 Location/Qualifiers
 1..1332
 /organism="Brassica oleracea"
 /mol_type="mRNA"
 /db_xref="taxon:3712"

gene
 1..1332
 /gene="CP5"

CDS
 32..1111
 /gene="CP5"
 /note="BoCP5; similar to Arabidopsis aleurain-like
 protease; expression increases as the broccoli heads
 senesce after harvest and during storage"
 /codon_start=1
 /product="senescence-associated cysteine protease"
 /protein_id="AAL60582.1"

ORIGIN
 48.4%; Score 518.8; DB 8; Length 1332;
 Query Match 68.6%; Pred. No. 7.2e-132; Indels 0; Gaps 0;
 Matches 715; Conservative 0; Mismatches 327;

QY 23 TCCTACTCTCTCTGCTCCGCCCTCTCCGTCGCGCTCCGCCGTTTCCGAAGAGCAATC 82
 Db 65 TTCTGCTGATCTCTCATCGCGCATCGACGCGGAGAGTATCGAATTCATGATGATCAAACC 125

QY 83 CAATCCGGTCCGTTACAAAGCCCTGACTCGATTGAGCCTCCATCCTCGCGCTCTTG 142
 Db 126 CGATCCGATATGGTCTCCGACCGCTCCCGGAGGTAGGAATCCCGTTGTCAGATCTTAG 185

QY 143 GCAGTTGCGGCCACGCTTCCACTTTCGCACGGTTCCGCCGAGGTAACGGAAGACTACG 202
 Db 186 GCGAATCAGCCACCGTATCTCTTCCGCTCGCTTCGCTCACAGGTACGGAAGAGGTACG 245
 QY 203 GATCGAGAGGAGGATCAAGAAAGAGTTTCGGGATCTTCGTGGAGAACTTAGGTTTATCC 262
 Db 246 AGAACGCGGAGGAGATGAAGCTCAGTTCTCGATATTCGAAGGAAATCTGGATTTCATCA 305
 QY 263 GGTCCACTAATCGGAAGGATCTGTGTATACCTTAGGAATCAACCAATTCGCGGACCTGA 322
 Db 306 GATCCACCAATAAGAAAGGCTTGTCTTCAAACTCGGTGTCATCAGTTTGTGTATGA 365
 QY 323 CTGCGAGGAATTCGCGACCAATCGCTTGTGTGCGCGCAGAACTGCTCGCGACTCGCC 382
 Db 366 CGTGGCAAGAGTTTCAAAGGACCAAGCTTGTGTGCTCAAAACTGCTCTGCCACTTTAA 425
 QY 383 ATGGAACACCGGTTTGTGATGGGTGCTTCTGTACAGAGGATTTGGAGGGGCAAG 442
 Db 426 AGGGAACCCACAACTCACAGGAGAGCTCTTCCGGAACCAAGACTCGGAGAGATG 485
 QY 443 GGATAGTGAGCCCTGTAAAGGACCAAGGAAGCTGTGGATCTTGTGGACTTTTCAGTACTA 502
 Db 486 GTATTGTTAGCCCTGTCAAAGACCAGGGGGTTGTGGATCGTGTGGACATTCAGTACTA 545
 QY 503 CTGAGCACTAGAGGCTGCATATACACACTAACTGGAAGAGACATCAATTCAGTAAC 562
 Db 546 CTGGAGCGCTTGAGGCAGCTTATCACCAAGCATTTTGGAAAAGGAATATCACTCTCTGAGC 605
 QY 563 AGCAACTTGTGACTGTGCTCAGCATTCATTAATCTTGTGATGCAATGAGGTTTGCCTT 622
 Db 606 AACAGCTTGTGATTTGTGTGGAGCTTTCACAACTATGCTGCAATGATGGCTTCCCTT 665
 QY 623 CCCAAGCCTTTGAATAGTTAAAGTACAATGGAGGATCGACACAGAACAGACTTATCCAT 682
 Db 666 CCCAAGCCTTTGAATACATCAATCCAGCGGCGCTCGACACGAGGAGCTTATCCTT 725
 QY 683 ACCTTGGTGCATGTTCTGCAACTTCAAGCAGGAGATGTTGTGTCAAGGTCAATG 742
 Db 726 ACACCGGTGAAGATGGAACCTGCAAGTATTCGTGTAACAGTCGCTGTAGAGTCTTGG 785
 QY 743 ATTTCGATAAACATCACCCCTGGGTGCTGAGGATGAGTTGAAGCATGAGTGGGTTGGTGC 802
 Db 786 ACTCAGTCAACATTTACTCTGGGTGCTGAAGATGAACGAAGCATGCGGTGATTTGGTAC 845
 QY 803 GTCCAGTTAGCTGCTATTGAGTTGTGAAAGTTTCAATCTGTACAAGAAAGGTAT 862
 Db 846 GACCAGTAAGCATAGCATTTGAGGTTTATACACTGCTCCGGCTTTTACAAGACGAGATT 905
 QY 863 ACAGCAGTGACACTGTGGAAGAGATCCCAATGATGTGAACCAACGAGTTCTTGGCGTCG 922
 Db 906 ACAGCGATAGTCACTGTGGACACTCCCAATGCAACGCTGTTTGGCGGTTG 965
 QY 923 GTTATGAGTCGAGGACGGGATTCCTTTATTTGGCTCATCAAGAACTCATGGGTACAAAT 982
 Db 966 GTTATGGAATTAAGACGGGTGTTCCATATTTGGCTTTAAAGAACTCATGGGGAGAGATT 1025
 QY 983 GGGGTGCAATGCTACTTTAAGATGGAACCTCGGCAAGAACATGTGTGTGTGTCGAACCT 1042
 Db 1026 GGGGTGCAAAAGTTTCTTCAAGATGGAATGCGGAGAACATGTGTGTGTTTGGACAT 1085
 QY 1043 GCGCATCTTATCCCATTTGGC 1064
 Db 1086 GTGCATCGTACCGGTTGTGGC 1107

RESULT 9
 PAU93166 1444 bp mRNA linear PLN 21-JAN-1998
 LOCUS
 DEFINITION Prunus armeniaca cysteine protease mRNA, complete cds.
 ACCESSION U93166
 VERSION U93166.1 GI:2677827
 KEYWORDS

```

SOURCE
ORGANISM
Prunus armeniaca (apricot)
Prunus armeniaca
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosids; eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
1 (bases 1 to 1444)
Mbeguile-A-Mbeguile, D., Gomez, R.-M. and Pils-Lycaon, B.
Sequence of AftPl, a Cysteine Proteinase from Apricot Fruit
(Accession No. U93166). Gene Expression During Fruit Ripening.
(PGR97-179)
JOURNAL
Plant Physiol. 115, 1730 (1997)
2 (bases 1 to 1444)
Mbeguile-A-Mbeguile, D., Gomez, R.-M. and Pils-Lycaon, B.
Direct Submission
Submitted (12-MAR-1997) Station de Technologie des Produits
Vegetaux, INRA, Site AGROPARC, Avignon 84914 Cedex 9, France
Location/Qualifiers
1. 1444
/organism="Prunus armeniaca"
/mol_type="mRNA"
/strain="Bergeron"
/db_xref="taxon:36596"
/clone="pAPR152"
/tissue type="mesocarp plus exocarp"
/dev stage="ripe fruit"
82..1158
/codon_start=1
/product="cysteine protease"
/protein_id="AAB97142.1"
/db_xref="GI:2677828"
/translation="MARVTLVLSAALVLAIVASGAAASSFDESNPIRLVSDGLRELEQ
QVVOVLGNRRALHPARFAHRYGKYESVEMKLYRIFSEKGLIRSTNKGKLPYTL
AVNFRADMSPEERFORLGAQNCASATTKGSHLTDVLPESKNWRBEGIVTPVQDG
HCGSWPESTGALRAYVOAFKQISLSKQOLVDCAGAFNFCGSLPQAFEXIK
YCGLDTEAAYPYGTGACKFSRNVGVQVLDVNTLGDRELKHAFAVFPVSVA
FOVVKSPRIYSGVYTTSDTSGSSPMNVNHAFLAVGYGEEGVFPWLIKNSWGESWDN
GYFKMEFGKNNCGVATCASYPDVA"

CDS
Query Match 48.3%; Score 517.2; DB 8; Length 1444;
Best Local Similarity 68.5%; Pred. No. 2e-131;
Matches 714; Conservative 0; Mismatches 328; Indels 0; Gaps 0;

QY 24 CCTACTCTCTCTGCTCTCGCCCTCTCGTCCGCTCGCGCTTCGAGAGAACATCC 83
DB 114 CTTGCTCTCTGCTGCGATCAGTGGCGGCGAGCAGCATCGAGTCCACACCC 173
QY 84 AATCCGCTCTCTACACAAAGCCCTGACTGATGAGCTGCCATCTCGCGCTCTGG 143
DB 174 CATCAGATTGCTATCGGATGCGCTCCGCTGATGAGCAGCAAGTCGCTCCGG 233
QY 144 CAGTTGCGCGCCACGCTTCCACTTCGACGCTTCGCGGAGTACGGGAGAGCTACGG 203
DB 234 CAACCTCTCGCGCGCTTCCACTTCGCTCGCGCTTCGCTCACCGTACGGGAGAGTATGA 293
QY 204 ATCGGAGAGAGATCAAGAGAGCTTCGGATCTTCGTGAGATCTAGCGTTATCCG 263
DB 294 GAGCGTGAGGAGATGAAGCTCGGTATGAGATTTCTCGGAGAAATGAAGCTGATCG 353
QY 264 GTCCACTAATCGGAAGGATCTGCTGTATACCTAGGAATCAACAAATCGCGACCTGAC 323
DB 354 ATCCACAACAGAGAGGCTTGCTTACCTTGTCTGCTGATCTGATGGAG 413
QY 324 CTGGAGGAGATTCGGGACCAATCGCTTGGTGGCGGAGAACTGCTCGGAGCTCGGCA 383
DB 414 TTGGGAAGAGTTTCAAGAGCAGAGTTGGAGCTGCCAGAACTGCTCGCCACCAAAA 473
QY 384 TGGAAACACCGGTTGTGTCGAGTGGCTGCTTCTGTAACAGGGGATTCGAGGAGCAAGG 443
DB 474 GGGAGCCACGAGCTCACTGATGCCGTTCTTCTGATCGAAAACACTGAGAGAGAGG 533
QY 444 GATAGTAGCCCTGTAAAGGACCAAGGAGCTGTGGATCTTCTGCGACTTTCACTATAC 503

```

ORIGIN

```

RESULT 10
AB032168 1340 bp mRNA linear PLN 24-JAN-2001
LOCUS Nicotiana tabacum NTCP-23 mRNA for cysteine protease, complete cds.
DEFINITION
AB032168
ACCESSION
AB032168.1 GI:8347419
VERSION
AB032168.1
KEYWORDS
cysteine protease.
SOURCE
Nicotiana tabacum (common tobacco)
ORGANISM
Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Nicotiana.
1 (sites)
Ueda, T., Seo, S., Ohashi, Y. and Hashimoto, J.
Circadian and senescence-enhanced expression of a tobacco cysteine
protease gene
JOURNAL
Plant Mol. Biol. 44 (5), 649-657 (2000)
MEDLINE
21039373
PUBMED
11198425
2 (bases 1 to 1340)
Ueda, T.
Direct Submission
Submitted (03-SEP-1999) Tadamasu Ueda, National Institute of
Agrobiological Resources, Department of Molecular Genetics; 2-1-2
Kannondai, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: uechu@ab.affrc.go.jp, Tel: 81-297-38-7443,
Fax: 81-298-38-7044)
Location/Qualifiers
1. 1340
/organism="Nicotiana tabacum"
FEATURES
source

```

```

DB 534 CATAGTGACCCCGATTAAAGACCAAGTCACTGTGGATCTTCTGGACATTTCAGCACCA 593
QY 504 TGGAGCATTAGAGGTGTCATATACACAGTAACTAGGAAGACACATATATCTGAACA 563
DB 594 TGGGGCTCTGAGGAGCTTATGTGAGGCAATTCGGAAGCAAAATCTCTCTCTGAGCA 653
QY 564 GCAACTTGGGACTGTGCTCAGCATTCATAAATTTGGATCAATGGAGGTTTGGCTTC 623
DB 654 CGAGCTTGGGATTTGCTGGAGCTTTCANAACTTTGGCTGCCATGGTGGTGGCTTGC 713
QY 624 CCAAGCCTTTGAATACGTTAAGTACAAATGGAGGCAATCGACACAGACAGATTATCCA 683
DB 714 TCAAGCTTTGAGTACATCAAAATACAAATGGCGGTCTCGACACCGAGGCTGCATATCT 773
QY 684 CTTTGGTGTCAATGTATCTGCAACTTCAAGCAGGAGATGTTGGTGTCAAGGTCAATTC 743
DB 774 CGTTGAAACGATGGGGCTTGCATAATTTTCAGCTGAAATGTTGGTGTCAAGTGTCTGA 833
QY 744 TTGATAAATACATCACCCCTGGGTGCTGAGGATGAGTTGAAGCATGAGTGGGCTTGGTGG 803
DB 834 CTCTGTCAATATCACCCCTGGGTGATGAACAAGAACTAAAGCATGCAAGTTGCGTTTGTGG 893
QY 804 TCAGTTAGCGTTGCAATTTGAGGTTGTGAAAGGTTTCAATCTGTACAGAAAGGTGTATA 863
DB 894 GCAGTCAAGCTGCGCTTTTCAGGTGGTCAAAAGCTTCGAAATTTTCAAGTCAGGAGTCTA 953
QY 864 CAGCAGTGACACCTGTGGAAGAGATCCAAATGATGTGAACCAACGAGTTCTTTCGCGCTCG 923
DB 954 CACCAGTGACCTTGTGGCAGCAGTCCCATGATGTGAACCATGCTGTCTTCTTCAGTTGG 1013
QY 924 TTATGAGTTCGAGGAGCGGATTCCTTATTTGGCTCATCAAGAACTCATGGGTACAAATTG 983
DB 1014 GTATGAGAGGAGGCGGCTGTCCTTCTGGCTCATCAAGAACTCTTGGGAGAGAGCTG 1073
QY 984 GGGTGCAATGGTACTTTAAGATGGAACCTCGCAAGAACATGTGTGTGTGTGCAACTTG 1043
DB 1074 GGGTGCAATGGTACTTTAAGATGGAAGTTGGAGAACATGTGTGTGTGTGCAACTG 1133
QY 1044 CGCATCTTATCCCATTTGGCT 1065
DB 1134 TGCATCGTACCCGATTTGTTGCT 1155

```

602	CTATCTGAGCAGCAGCTTGTGGACTGTCTGCGAGCTTTTAAATACTTTGGCTGCAATGGT	661
613	GGTTTGGCTTCCCAAGCCTTTGAATACGTTAAGTCAATGGAGGCATCAGACAGAACAG	672
662	GGGCTCCATCACAAGCCTTTGAGTATATTAAATCCAAATGGTGGTCTTGACACTGAAGAA	721
673	ACTTATCCATACCTTGGTGTCAATGGTATCTGCAACTTCAAGCAGGAGAAATGTTGGTGT	732
722	GCATATCCATACACTGCGAAGATGGCTTATGTAATTTCTCATCAGAAAAATGTTGGTGT	781
733	AAGGTCAATGATTCGATAAATCACCCTGGGTCTGAGATGAGTGAAGCATGAGTGAAG	792
782	AAAGTCATCATTCGCTCAATATATACCTGGGTGCTGAAGATGAACTAAATACGCGGT	841
793	GGCTTGGTGGCTCCAGTTCAGTTAGCTTTGAGTTTGAAGGTTTCAATCTGTACAAG	852
842	GCATTTGGTAGGCCCTTTAGTATAGCTTTTGGAGTGATAAAGGTTTCAACAAATCAAG	901
853	AAAGGTGTATACAGCATGACCTGTGGAAGAGATCCATGATGTAACCAACGAGTT	912
902	AGTGGTGTATACACAGCAGCAATGCGGCAACCACTCCCATGGATGTAAACCAATGCT	961
913	CTTGGCTGGTATGAGTCGAGGACGGATTCCTTATTTGGTCTCATCAAGAACTCATGG	972
962	CTTGGTGGTGGTACGGTGTGAAATGGTGTTCCTTATTTGGTCTCATCAAGAACTCATGG	1021
973	GGTACAAATTCGGGTGACAAATGGCTACTTTAAGATGGAATCGGCAAGACATGTGTGGT	1032
1022	GGAGCAGATTGGGTGACAAATGGATATCTTCAAAATGGAGATGGGAAAGACATGTGTGGT	1081
1033	GTTCGAATTCGGCATCTTATCCCATTTGGC	1064
1082	ATTGGCACTTGGCATCTTACCTCTGCTGTGGC	1113
RESULT 11		
AF233883	1305 bp mRNA linear	PLN 17-JUL-2000
LOCUS	Arabidopsis thaliana AALP protein mRNA, complete cds.	
DEFINITION	AF233883	
ACCESSION	AF233883.1	GI:7230639
VERSION		
KEYWORDS	Arabidopsis thaliana (thale cress)	
SOURCE	Arabidopsis thaliana	
ORGANISM	Arabidopsis thaliana	
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.	
AUTHORS	1 (bases 1 to 1305) Ahmed, S.U., Rojo, E., Kovaleva, V., Venkataraman, S., Dombrowski, J.E., Matsuoka, K., and Raike, N.V.	
TITLE	The plant vacuolar sorting receptor AtELP is involved in transport of NH(2)-terminal propeptide-containing vacuolar proteins in Arabidopsis thaliana	
JOURNAL	J. Cell Biol. 149 (7), 1335-1344 (2000)	
MEDLINE	20330156	
PUBMED	10871276	
REFERENCE	2 (bases 1 to 1305)	
AUTHORS	Raikhel, N.	
TITLE	Direct Submission	
JOURNAL	Submitted (13-FEB-2000) MSU-DOE Plant Research Lab, MSU, Plant Biology Building, East Lansing, MI 48824, USA	
FEATURES	Location/Qualifiers	
source	1..1305	
	/organism="Arabidopsis thaliana"	
	/mol_type="mRNA"	
	/db_xref="taxon:3702"	
	10..1086	
	/note="similar to aleurain"	
	/codon_start=1	
	/product="AALP protein"	
	/protein_id="AAF43041.1"	
	/db_xref="GI:7230640"	
CDS		

/translation="MSAKTILSSVVLVLAASAAANIGFDESPIRMVSDGLREVEE
SVSQILQSRHVLSPARFTHRYKQYQNVVEEMKLRFSIPKENLDLRSTNKKGLSVKL
GVNQFADLTQWERTKLGAAQNSATLKGSHVTEAALFETKDWREDEIVSPFRDQG
CGCSWTFSTGSALEAAHQAFKGISLSEQQLVDCAGAFNNYNGNGLPSQAFXYIK
SNGGLDPEKAYPTGKDETKFSAENVGVQLNSVNITLGADELKHLVGLVRPVSIA
FEVHSFRLYKSGVYTDSHCGSTPMDVNHAVLAVGVGEDVGPYMLIKNSWGDWGDK
GYFWMENKMKCGIATCASYPVVA"

ORIGIN

Query Match 45.6%; Score 488.4; DB 8; Length 1305;
Best Local Similarity 67.0%; Pred. No. 1.8e-123;
Matches 693; Conservative 0; Mismatches 341; Indels 0; Gaps 0;

QY	32	TCCTGGTCTCGCCCTCTCGTCCGCTCGCGGTTTCGGAAGACAAATCCCAATCCGGT	91
DB	50	TTCTCGTCGCGCATCAGCAGCGCTAATATCGATTGATGAGTCAAAACCGATCCGAA	109
QY	92	CCGTTACACAAGGCCGTGACTCGATTGAGCGTCGCCATCTCTCGGCGTCTTGGCAGTTGCC	151
DB	110	TGGTCTCCGATGGTCTCGGAGGTAGAGAATCTGTTTCCAGATCTTAGGTCAATCTC	169
QY	152	GCCAGCGCTTCCACTTCGACGGTTCGCCGAGTACGGAAGAGCTACGGATCGGAGG	211
DB	170	GTCAGTCTCTCTTTCGTCGCTTCACTCACCAGATGTTAAAGTATCAGAACGTGG	229
QY	212	AGGAGATCAAGAGAGGTTCCGGATCTTCGTGGAGAACTAGCGTCTTATCCGTTCCACTA	271
DB	230	AGGAGATGAAGCTTCGATTCTCGATTTTCAAGGAGATCTTGATTTGATCAGATCCACCA	289
QY	272	ATCGGAGGATCTCTGATATACCTTAGAATCAACCAATTCGCGACCTCAGCTGGGAGG	331
DB	290	ACAAGAAAGGTTATCTTACAAATCGGTTGTAATCAATTTGCTGATTTGACATGGCAAG	349
QY	332	AATTCGCGACCAATCGCCTTGGTCGCGCGCAGAACTCTCGGCGACTCGCATGGAAACC	391
DB	350	AGTTTCAAGACCAAGCTTGCTGCTCAGAACTGCTCTGCCACTTTAAAGGGCAGCC	409
QY	392	ACCGGTTTGTGATGGCGTGCTTCTGTACAGGAGGATTCGAGGAGCAAGGATAGTGA	451
DB	410	ACAAGGTACAGAAAGCAGCTCTTCTGTAACAAAGAACTGGAGAGAGATGGTATCGTTA	469
QY	452	GCCCTGTAAGAGCAACAGGAAGCTGTGATCTTCTCGGACTTTTCAGTACTCTGAGCAC	511
DB	470	GTCCGGTCAAGATCAGGAGGTTGTGATCTTCTGAGCAATTCAGCAAACTGGAGCTC	529
QY	512	TAGAGGTGCATATACACAGTAACTGGAAGAGCAATCATATCTGAAACAGCAACTTG	571
DB	530	TTGAGGAGCTTACCATCAGGCATTTGGAAGAGATATCTCTCTGAGCAACAGCTTG	589
QY	572	TGGACTGTGCTCAGCATTCATTAACCTTGGATGCAATGAGGTTTCCCTTCCCAAGCCT	631
DB	590	TGGATTGTGCTGGAGCTTTCATTAACCTATGTTGCAATGTGGGCTTCTTCTCAAGCCT	649
QY	632	TTGAATAGTTTAACTACAAATGAGGAGCTCGACAGAACAGACTTATCCATACCTGGTG	691
DB	650	TTGAATACATAAATCCAAAGGTGGCTCGACAGACAGAGAAAGCTTATCCTTATACCGGTA	709
QY	692	TCAATGTATCTGCAACTTCAAGCAGGAGATGTTGTGTCAAGTCAATGATTCGATAA	751
DB	710	AAGATGAACCTGCAAAATTTTCAGCTGAAACGTTGTGTGTAAGAGTCTCAACTCAGTCA	769
QY	752	ACATCACCTCGGTGCTGAGGATGATGTAAGCATGAGTGGGCTTGGTGGCTCCAGTTA	811
DB	770	ACATTAATCTGGTGTGTAAGATGAATGTAAGCATGCGTTGGATTGTACGGCCAGTAA	829
QY	812	CGGTGCAATTTGAGTTGTGAAGGTTTCAATCTGTACAGAAAGGTTATACAGCAGTG	871
DB	830	GCATAGCATTTGAGGTTATACACTCGTTCCGGCTTTTCAAGAGTGGAGTTTACACTGATA	889
QY	872	ACACTGTGGAAGAGATCCATGATGTGAACACGAGGTTCTTCCGCTCGGTTATGGAG	931
DB	890	GTCAGTGTGGAAGTACTTCAATGATGTAAGAACCAACGCGGTTTGGCCGTTGGTTATCGAG	949

QY	932	TCGAGACGGGATTCCTTATTGGCTCATCAAGAACTCATGGGTACAAATTTGGGCTGACA	991
DB	950	TTGAAGACGGTGTACCATATTGGCTTATTAAGAACTCATGGGACGGATTGGGCGGACA	1009
QY	992	ATGGCTACTTTAAGATGAACATCGCAAGAACATGTGGTGTTCGCAACTTGGCGCATCTT	1051
DB	1010	AAGGTTACTTCAAGATGAGATGGGGAAGAACATGTGTGTTGCTACATGTGCATCAT	1069
QY	1052	ATCCCATTTGGCT	1065
DB	1070	ACCCGTTGTGGCT	1083

RESULT 12

LOCUS AY088662 1352 bp mRNA linear PLN 14-APR-2003
DEFINITION Arabidopsis thaliana clone 8989 mRNA, complete sequence.
ACCESSION AY088662
VERSION AY088662.1 GI:21407436
KEYWORDS FLI CDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

HAAS,B.J., VOJFOSKY,N., TOWN,C.D., TROUKHAN,M., ALEXANDROV,N.,
FELDMANN,K.A., FLAVELL,R.B., WHITE,O. and SALZBERG,S.I.
Full-length messenger RNA sequences greatly improve genome

annotation

Genome Biol. 3 (6), RESEARCH0029 (2002)

JOURNAL MEDLINE

PUBMED 22088475

REFERENCE 2 (bases 1 to 1352)

Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and

Feldmann,K.

Full-length cDNA from Arabidopsis thaliana

Unpublished to 1352)

Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and

Feldmann,K.

Direct Submission

Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road,

Malibu, CA 90265, USA

This clone sequence is one of 5,000 Ceres full-length cDNAs made

available to fRGR and Genbank. The following quality assessment of

this set was done by comparison with known proteins: two percent of

the clones are estimated to be 5'-truncated; less than one percent

are 3'-truncated; approximately two percent represent alternative

splice variants, including unspliced introns and spliced exons; one

percent may contain premature stop codons; five percent may have

frame shifts in a coding region. A sequence is considered to be

5'-truncated if it lacks the translation initiation start (ATG). A

sequence is considered to be 3'-truncated if it lacks the

C-terminal end of the encoded protein. Please note that these cDNA

sequences are derived from the Ws or Laer ecotypes and therefore

may contain polymorphisms when compared to sequences from Col-0.

Genet carried out the library production and sequencing of the

full-length clones. Ceres, Inc. carried out the clustering of the

5' sequences, selection of clones, and sequence assembly.

Location/Qualifiers

1. .1352

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/clone="8989"

80..1156

/codon_start=1

/product="cysteine proteinase AAP"

/protein_id="AA066984.1"

/db_xref="GI:21617934"

/translation="MSAKTILSSVVLVLAASAAANIGFDESPIRMVSDGLREVEE

SVSQILQSRHVLSPARFTHRYKQYQNVVEEMKLRFSIPKENLDLRSTNKKGLSVKL

GVNQFADLTQWERTKLGAAQNSATLKGSHVTEAALFETKDWREDEIVSPFRDQG

CGCSWTFSTGSALEAAHQAFKGISLSEQQLVDCAGAFNNYNGNGLPSQAFXYIK

SNGGLDPEKAYPTGKDETKFSAENVGVQLNSVNITLGADELKHLVGLVRPVSIA

FEVHSFRLYKSGVYTDSHCGSTPMDVNHAVLAVGVGEDVGPYMLIKNSWGDWGDK

GYFWMENKMKCGIATCASYPVVA"

SVSQILQSRHVLSPARFTHRYKQYQNVVEEMKLRFSIPKENLDLRSTNKKGLSVKL

FEATURES

source

CDS

GVNPFADLTWQEFORTKLGAAQNCATLKGSHKVTBAALPETKMDREBDGIVSPVQDG
CGSCWTFSTGALERAYHQAQKGLISLSEQLVDCAAFNNYGNGLPSQAFYIK
SNGGLDTEKATPTGKDETKFSAENVGVQVLSVNTLGAEBELKHAVALRPSVIA
FEVHSFRLYKSGVYTDHSCGSTPMVNHAVLAVGVGVEDGVFWLILKNSWADWGDK
GYFKWEMGNKMGCIATCASYPVVA"

ORIGIN

Query Match 45.6%; Score 488.4; DB 8; Length 1352;
Best Local Similarity 67.0%; Pred. No. 1.8e-123; Indels 0; Gaps 0;
Matches 693; Conservative 0; Mismatches 341;

QY 32 TCCTGGTCTCCGCCCTCTCCGTCGCGCTCGCGGTTTCGAAGAAGACAAATCCAAATCCGGT 91
DB 120 TTCTCGTCCGCGCATCAGCAGCGGCTAATATCGGATTCGATGAGTCAAACCCGATCCGAA 179
QY 92 CGGTTACAAAGGCTGACTCGATTGAGCTGCCATCTCGGCGTCTTCGGCAGTTGCC 151
DB 180 TGGTCTCCGATGGTCTCCGGAGGTAGAGAAATCTGTTTCCAGATCTTAGTCAATCTC 239
QY 152 GCCAGCCTTCCACTTCGCACGGTTTCGCCCGCAGGATCGGAAGAGCTACGGATCGGAGG 211
DB 240 GTACGTTCTCTCTTCTGCTGCTTCACTCACCAGATATGTAAGAGTATCAGAACGTTGG 299
QY 212 AGGAGATCAAGAGAGTTCGGATCTTCGTGGAGAACTAGCGTTTATCCGTCCTCACTA 271
DB 300 AGGAGTGAAGTTCGATTTCTCGATTTTCAAGGAGATCTTGATTTGATGATGATCCACCA 359
QY 272 ATCGGAAGGATCTGCTGATATACCTTAGGAATCAACCAATTCGCCACCTGACCTGGGAGG 331
DB 360 ACAAGAAAGGCTTATCTTACAAACTCGGTGTTTAAATCAATTTGCTGATTTGATGATGCAAG 419
QY 332 AATTCGGACCAATCGCCTTGTGTCGGCGCAGAACTGCTCGCGACTCGGCATGGAACC 391
DB 420 AGTTTCAAAGGACCAAGCTTGTGCTGCTCAGAACTGCTCTGCCACTTTAAAGGCGAGCC 479
QY 392 ACGGTTTCTGATGCGGTCTTCTGTAAACAGGAGTTGGAGGAGCAAGGATAGTGA 451
DB 480 ACAAGTCAAGAGCAGCTCTTCTGAAACAAAGAGCTGGAGAGAGATGATCGTTA 539
QY 452 GCGCTTAAAGGACCAAGAGCTGTGATCTTCTGAGATCTTCTGAGTCTTCTGAGTCTGAGCAG 511
DB 540 GTCCGCTCAAAGATCAGGAGGTTTGGATCTTGTGGACATTCAGACAACTGAGGCTC 599
QY 512 TAGAGCTCATATACAGCTTAACCTGGAAGAGCAGATCAATCTCTGAACCAAGCAACTTG 571
DB 600 TTGAGGAGCTTACCATCAGGCAATTTGGAAGAGATATCTCTCTGAGCAAGAGCTTG 659
QY 572 TGGACTGTGCTCAGCAATCAATACTTTGGATGCAATGAGGTTTGCCTTCCCAAGCT 631
DB 660 TGGATTGTGCTGAGCTTTCAATACTATGTTGCAATGCTGCGCTTCTCTCAAGCT 719
QY 632 TTGAATACGTTAAGTACATGAGGAGCTGACACAGACAGACTTATCCATACCTTGGTG 691
DB 720 TTGAATACATAAATCAAGCGTGGCTGACACAGAGAAAGCTTATCTTATACCGTGA 779
QY 692 TCAATGATCTGCAACTTCAAGCAGGAGAAATTTGGTGTCAAGGTCAATGATTCGATAA 751
DB 780 AAGTGAACCTGCAAAATTTTACGCTGAAACGTTGGTGTACAAAGTCTCAACTCAGTCA 839
QY 752 ACATACCCCTGGGTGCTGAGATGAGTTGAGCATGAGTGGCTTGGTGGCTTCCAGTTA 811
DB 840 ACATATCTCGGTGCTGAAGATGAATGAAGCATGCGGTGGAATGGTACGCGCCAGTAA 899
QY 812 GCGTTGATTTGAGTTGTGAAGGTTTCAATCTGTACAGAAAGGTGTATACAGAGTG 871
DB 900 GCATAGATTTGAGTTTATACATCTGTTCCGGCTTTTACAGAGTGGAGTTTACACTGATA 959
QY 872 ACACCTGTGAAGAGATCCAAATGATGTAACACGAGGTTCTTGGCTGGTTATGGAG 931
DB 960 GTCACTGTGAAGTACTCAATGATGTGAACACGCGGTTTGGCGGTTGGTTATGGAG 1019
QY 932 TCGAGGACGGATTCCTTATTTGGCTCATCAAGAACTCATCGGTTACAAATTTGGGTTGACA 991

DB 1020 TTGAAGACGGTGTACATATTCGCTTATTAGAACTCATGGGAGCGGATTCGGCGGACA 1079
QY 992 ATGGCTACTTTAAGATGGAATCGGCAAGAACATGTGTGGTGTTCAACTTCGCGCATCTT 1051
DB 1080 AAGGTTACTTCAAGATGGAGATGGGGAAGAACATGTGTGGTATTGCTATCATGTGATCAT 1139
QY 1052 ATCCCATTTGGCT 1065
DB 1140 ACCCGTTTGGCT 1153

RESULT 13

AF083703 LOCUS 1355 bp mRNA linear PLN 29-OCT-2002
DEFINITION Arabidopsis thaliana clone sps232 unknown mRNA.
ACCESSION AF083703
VERSION AF083703.1 GI:24417303

KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana

Subaryota; viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 1355)
AUTHORS Stracke, R. and Palme, K.
TITLE Signal Peptide Selection derived cDNAs from Arabidopsis thaliana
leaves and guard cells

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1355)

AUTHORS Stracke, R. and Palme, K.

TITLE Direct Submission

JOURNAL Submitted (12-AUG-1998) Max-Planck-Laboratorium in der

Max-Planck-Gesellschaft, Carl-von-Linne-Weg 10, Koeln D-50829,

Germany

FEATURES Location/Qualifiers

1..1355

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/clone="sps232"

/notes="signal peptide selection derived"

66..1142

/codon_start=1

/product="unknown"

/protein_id="AA060262.1"

/db_xref="GI:24417304"

/translation="MSAKTILSSVWLVLVAASAAANIGPDSNPIRMVSDGLREVEE
SVQILGOSRHVLSFAFTHRYGKYKNVEEMKLRFSI FKENLDLIRSTNKGILSYKL
GVNQFADLTWQEFORTKLGAAQNCATLKGSHKVTBAALPETKMDREBDGIVSPVQDG
CGSCWTFSTGALERAYHQAQKGLISLSEQLVDCAAFNNYGNGLPSQAFYIK
SNGGLDTEKATPTGKDETKFSAENVGVQVLSVNTLGAEBELKHAVALRPSVIA
FEVHSFRLYKSGVYTDHSCGSTPMVNHAVLAVGVGVEDGVFWLILKNSWADWGDK
GYFKWEMGNKMGCIATCASYPVVA"

ORIGIN

Query Match 45.6%; Score 488.4; DB 8; Length 1355;
Best Local Similarity 67.0%; Pred. No. 1.8e-123; Indels 0; Gaps 0;
Matches 693; Conservative 0; Mismatches 341;

QY 32 TCCTGGTCTCCGCCCTCTCCGTCGCGCTCGCGGTTTCGAAGAAGACAAATCCAAATCCGGT 91
DB 106 TTCTCGTCCGCGCATCAGCAGCGGCTAATATCGGATTCGATGAGTCAAACCCGATCCGAA 165
QY 92 CGGTTACAAAGGCTGACTCGATTGAGCTGCGCTGCCATCTCGGCTCTCTCGCAGTTGCC 151
DB 166 TGGTCTCCGATGGTCTCCGGAGGTAGAGAAATCTGTTTCCAGATCTTAGTCAATCTC 225
QY 152 GCCAGCCTTCCACTTCGCACGGTTTCGCCCGCAGGATCGGAAGAGCTACGGATCGGAGG 211
DB 226 GTCACTGTGAGTACTCAATGATGTGAACACGCGGTTTGGCGGTTGGTTATGGAG 285
QY 212 AGGAGATCAAGAGAGGTTCCGGATCTTCGTGGAGAACTCATCGGTTTATCCGTTCCACTA 271

Db	286	AGGAGATGAAGCTTCGATTCGATTTTCAAGGAGAACTTGTGATTCAGATCCACCA	345	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Qy	272	ATCGAAGAGATCTGTGTATACCTAGGAATCAACCAATTCGCGACCTGACTGGAGG	331	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Db	346	ACAAGAAAGGCTTATCTTCAAACTCGGTGTTAATCAATTTGATTCGATGCAAG	405	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Qy	332	AATTCGGACCAATCGCTTGTGCGGCGCAGAACTGCTCGGCACTGGCGATGAAACC	391	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Db	406	AGTTTCAAGGACCAAGCTTGTGCTGCTCAGAACTGCTCGCCACTTTAAAGGCGAGC	465	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Qy	392	ACCGTTTGTGATGGGCTGCTTCCTGTAAAGAGGATTTGGAGGAGCAAGGATAGTGA	451	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Db	466	ACAAGGTACAGAAGCAGCTCTTCTCCTGAACAAAGACCTGGAGAGATGATTCGTTA	525	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Qy	452	GCCTGTAAAGGACCAAGGAGCTGTGATCTTCTGCTGCACTTTTCAGTACTACTGGAGCAC	511	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Db	526	GTCCGTTCAAGATCAGGAGGTTTGATCTTGCTGGCAATTCACACAACTGGAGCTC	585	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Qy	512	TAGAGCTGCAATPACACAGCTAACTGGAAGAGCACATCATTTCTGAACAGCACTTG	571	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Db	586	TTGAGGCGCTTACCATCAGGCATTTGGAAAGGAATATCTCTCTGAGCAACAGCTTG	645	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Qy	572	TGGACTGTGCTCAGCATTCATTAATCTTTGGATGCAATGGAGTTTGCCTTCCCAAGCCT	631	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Db	646	TGGATTGTCTGGAGCTTCAATAACTATGTTGCAATGGTGGCTTCTCTCAAGCCT	705	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Qy	632	TTCAATACGTTAAGTACAATGAGGAGCATCGACACAGACAGCTTATCCATACCTTGGTG	691	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Db	706	TTGNATACATAAATCAACGGTGGCTCGACACAGAGAAAGCTTATCTTATACCGGTA	765	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Qy	692	TCATGTGATCTGCAACTCAAGCAGGAGAAATGTTGGTGTCAAGTCAATGATTGATTA	751	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Db	766	ARGATGAACCTGCAATTTTACGTGAAACAGTTGGTGACAAAGTCTCAACTCAGTCA	825	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Qy	752	ACATCACCTCGGTGTGAGGATGAGTTGAAGCATGCAGTGGCTTGTGCGTCCAGTTA	811	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Db	826	ACATTACTCTGGTGTGTAAGTGAATGAAGCATGCGTGGTGGTACGCGCCAGTAA	885	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Qy	812	CGGTTCATTTGAGTTGTGAAGGTTTCAATCTGTACAAAGAGTGTATACAGCAGTG	871	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Db	886	GCATAGCATTTGAGTTATACATCTGTTCCGGCTTTTACAAGAGTGAGTTTACACTGA	945	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Qy	872	ACACCTGTGAAGAGATCCCAATGATGTGAACACGAGTCTTCCGCTCGGTTATGGAG	931	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Db	946	GTCACTGTGAAGTACTCCATGATGTGAACACGCGGTTTGGCGGTGTTATGGAG	1005	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Qy	932	TCGAGGACGGATTCCTTATTGGCTCATCAAGAACTCATGGGTACAAATGGGGTGACA	991	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Db	1006	TTGAAGACGGTGTACCATATTGGCTTATTAAGAACTCATGGGAGCGGATTGGGGCGACA	1065	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Qy	992	ATGGCTACTTTAGATGGAATCGGCAAGACATGTGTGTTGCAACTTGGCATCTT	1051	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Db	1066	ARGGTTACTTCAAGATGGAGTGGGAAGAACATGTGTGTTATGTTACTATGTCATCAT	1125	Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Qy	1052	ATCCCATTTGGCT 1065		Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Db	1126	ACCCCGTTGGCT 1139		Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
RESULT 14				
BT000676		1331 bp	mRNA	linear
LOCUS				PLN 01-OCT-2002
DEFINITION				Arabidopsis thaliana clone RAFLO8-12-G17 (R11172) putative cysteine proteinase AALP (At5g60360) mRNA, complete cds.
ACCESSION				BT000676
VERSION				BT000676.1
KEYWORDS				GI:23397073
SOURCE				FLI CDNA.
ORGANISM				Arabidopsis thaliana (thale cress)
				Arabidopsis thaliana
				Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
				Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
				rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE				1 (bases 1 to 1331)

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

gene

5'UTR

misc_difference

2..4

CDS

3'UTR

Db 1069 AAGGTTACTTCAAGATGGAGATGGCGAAGAACATGTGTGTTGTTACATGTGCATCAT 112

Qy 1052 ATCCCAATTGGGCT 1065

Db 1129 ACCCGTGTGGCT 1142

RESULT 15

AF360273

LOCUS Arabidopsis thaliana putative cysteine proteinase AAP (At5g60360)

DEFINITION mRNA, complete cds.

ACCESSION AF360273

VERSION AF360273.1 GI:13430721

KEYWORDS FLI CDNA.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 1370)

AUTHORS Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.

TITLE Arabidopsis Full Length cDNA Clones

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1370)

AUTHORS Yamada,K., Liu,S.X., Sakano,H., Phan,P.K., Banh,J., Chung,M.K., Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.

TITLE Direct Submission

JOURNAL Submitted (13-MAR-2001) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA

COMMENT RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL CDNA : 'RIKEN Arabidopsis Full-length cDNA') : Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PCEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Yamada,K., Liu,S.X., Sakano,H., Phan,P.K., Banh,J., Chung,M.K., Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L., Chen,H., Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Lam,B., Lin,J., Miranda,M., Nguyen,M., Palm,C.J., Shinn,P., Southwick,A., Davis,R.W., Ecker,J.R. and Theologis,A.

Yamada,K. (SSP/PCEC) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PCEC) contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis genome submitted to Genbank.

FEATURES

source

1. 1370

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/chromosome="5"

/clone="RAFL09-18-G19 (R09257)"

/ecotype="Columbia"

/note="This clone is in a modified pBluescript vector (FLC-1) as a BamHI/XhoI insert."

1. 1370

/gene="At5g60360"

5'UTR	1. .39	/gene="At5g60360"	
	misc_difference 1. .2	/note="compared to genomic sequence"	
		/replace="tg"	
	misc_difference 4		
CDS	/gene="At5g60360"	/note="compared to genomic sequence"	
	/replace="g"		
	40. .1116		
	/gene="At5g60360"	/note="contains a papain family cysteine protease domain PF00112"	
3'UTR	/codon_start=1		
	/evidence=experimental		
	/product="putative cysteine proteinase AAP"		
	/protein_id="AAK25983.1"		
misc_feature	/db_xref="GI:13430722"		
	/translation="MSAKTILSSVLVLAASAAANIGFDESNPIMVSDGLREVEE SVQILQGSRRHLVSFAFTHRYGKYQNVSEMKLRSIFKENLDLIRSNKGLSKL GVNQFADLTWQEFRTKLGAAQNCSTLKGSHKVFTEALPDEKDWREDGIVSVQDG GCSCWTFSTTGALEAYHOAFQKISLSEQQLVDCAGAFNNGCNGLPSPQAFYIK SNGGLDEKAVPYTKDETCKFSAENVGVOLNSVNTTLGAEDLKHAVGLVRPVSIA FEVTHSPRLYKSGVYTDHSGSTPMDVNHAVLAVGVEDGVFWLKNWSGADWDK GYFKMENGKMKGIATCASYPVA"		
	1117. .1370		
	/gene="At5g60360"		
misc_feature	1354		
	/gene="At5g60360"		
	/note="artifact within poly A tail"		

ORIGIN

Query Match 45.5%; Score 486.8; DB 8; Length 1370;
Best Local Similarity 66.9%; Pred. No. 5e-123;
Matches 692; Conservative 0; Mismatches 342; Indels 0; Gaps 0;

Qy	32	TCCTGGTCTCGCCCTCTCGTCCGCTCGCGGTTTCGAAGAAGACAAATCCATCCGGT	91
Db	80	TTCTGTCGCCGATCAGCAGCGCTAATTCGATTCGATGATCAAAACCGATCGAA	139
Qy	92	CCGTTACACAAAGCCCTGACTCGATTGAGCTGCCATCTCGCGGCTCTTGCGCATTTGCC	151
Db	140	TGGTCTCGATGGTCTCGCGGAGGTAGAGAATCTGTTCCAGATCTTAGGTCAATCTC	199
Qy	152	GCCAGCCTTCCATCTCGCAGGTTCCGCCGAGGTACGGAGCAGCTACGGATCGGAGG	211
Db	200	GTACAGTTCTCTCTCTGCTCGCTTCACTACCGATATGGTAAAGATATCAGAACGTGG	259
Qy	212	AGGAGATCAAGAAGAGGTTCCGGATCTTCGTGGAGATCTAGCGTTTATCGGTCCACTA	271
Db	260	AGGAGATGAAGCTTCGATTTTCGATTTTCAAGGAGATCTTGAATTCAGATCCACCA	319
Qy	272	ATCGAAGGATCTGTGTAATACCTAGGAATCAACCAATTCGCCGACCTGACCTGGGAGG	331
Db	320	ACAAGAAAGGCTTATCTTACAAATCCGGTGTAAATCAATTTGCTGATTTGACATGGCAAG	379
Qy	332	AATTCGGACCAATCGCTTGGTGGCGGACAGAACTGCTCGCGCACTCGCATCGAAACC	391
Db	380	AGTTTCAAGAGCAAGCTTGGTGTCTCAGACTGCTCTGCCACTTAAAGGGCAGCC	439
Qy	392	ACCGGTTGTGATGGCGTGTCTCTGTAAAGAGGATGGAGGAGCAAGGATAGTGA	451
Db	440	ACAAGGTACAGAAAGCAGCTCTTCTGAAACAAAGACTGGAGAGAAAGATGGTATCGTTA	499
Qy	452	GCCCTGTAAGAGCAAGGAGCTGTGGATCTTGTGGACTTTCAGTACTCTGAGCAGC	511
Db	500	GTCCGGTCAAGATCAGGAGGTTGTGGATCTTGTGGACATTCAGCACACTGGAGCTC	559
Qy	512	TAGAGGCTGCATATACAGCTTAATCTGAAAGAGCACATCAATATCTGAAACAGCAATTG	571
Db	560	TTGAGGCGAGCTTACCATCAGGCATTTGGAAAGGAATTTCTCTCTCTGAGCAACAGCTTG	619

Search completed: March 29, 2004, 20:28:32
Job time : 2968 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 17:13:57 ; Search time 373 Seconds
(without alignments)
12197.909 Million cell updates/sec

Title: US-10-087-714-1
Perfect score: 1071
Sequence: 1 atggcagctaaagtctctt.....atccattgtggtgtgtag 1071

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1071	100.0	1071	9	Adc61376 DNA encod
2	569.6	53.2	1442	2	Rat35111 Enhanced
3	541.4	50.6	1407	4	Aaf44802 Cysteine
4	541.4	50.6	1407	9	Adb94743 Programme
5	521.6	48.7	1467	2	Av82457 Triticum
6	488.4	45.6	1358	3	Aac37877 Arabidops
7	486.8	45.5	1370	3	Aac48945 Arabidops
8	478.2	44.6	1363	3	Aac40628 Arabidops
9	475.8	44.4	1388	2	Rat35100 Up-regula
10	467.6	43.7	1077	3	Aac51435 Arabidops
11	467.6	43.7	1282	6	Abn98252 Arabidops
12	429.8	40.1	1072	3	Aac44535 Zea mays
13	429.4	40.1	2149	3	Aac44178 Arabidops
14	407.4	38.0	1652	9	Adb94877 Programme
15	351.8	32.8	888	9	Adb94874 Programme
16	349.8	32.7	675	8	ACL19836 DNA clone
17	348.8	32.6	618	8	ACL19838 DNA clone
18	348.8	32.6	662	8	ACL19837 DNA clone
19	327.6	30.6	590	8	ACL19847 DNA clone
20	327.2	30.6	630	8	ACL19848 DNA clone
21	323.4	30.2	565	8	ACL19840 DNA clone
22	321.6	30.0	644	8	ACL19841 DNA clone
23	305.8	28.6	604	8	ACL19843 DNA clone

C 24	304	28.4	536	8	ACL19835	ACL19835 DNA clone
C 25	298.4	27.9	588	7	ABX56737	ABX56737 Arabidops
C 26	297.8	27.8	595	8	ACL19839	ACL19839 DNA clone
C 27	293	27.4	578	8	ACL19845	ACL19845 DNA clone
C 28	285.4	26.6	550	8	ACL19846	ACL19846 DNA clone
C 29	283.4	26.5	545	8	ACL19849	ACL19849 DNA clone
C 30	280.2	26.2	677	4	Aaf44806	Aaf44806 Cysteine
C 31	280.2	26.2	677	9	ADB94747	ADB94747 Programme
C 32	278.2	26.0	499	8	ACL19831	ACL19831 DNA clone
C 33	269.8	25.2	1362	7	ABT41873	ABT41873 Toxicity
C 34	269.8	25.2	1362	9	ADB52643	ADB52643 Primary r
C 35	257.6	24.1	714	8	ACL19871	ACL19871 DNA clone
C 36	253.2	23.6	530	8	ACL19827	ACL19827 DNA clone
C 37	249.4	23.3	609	8	ACL19854	ACL19854 DNA clone
C 38	249.4	23.3	710	8	ACL19865	ACL19865 DNA clone
C 39	248.6	23.2	718	8	ACL19870	ACL19870 DNA clone
C 40	247.2	23.1	723	8	ACL19866	ACL19866 DNA clone
C 41	244.4	22.8	1399	6	ABK84545	ABK84545 Human cDN
C 42	244.4	22.8	1433	9	ADD18512	ADD18512 Human pro
C 43	244	22.8	1475	4	AAI93479	AAI93479 Human pol
C 44	244	22.8	1475	4	AAI93479	AAI93479 Human pol
C 45	244	22.8	1475	4	AAI93494	AAI93494 Human pol

ALIGNMENTS

RESULT 1
ADC61376
ID ADC61376 standard; DNA; 1071 BP.
AC ADC61376;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA encoding Vanilla planifolia 4-hydroxybenzaldehyde synthase (4HBS).
XX
KW Vanilla production; plant; vanillin biosynthesis;
XX 4-hydroxybenzaldehyde synthase; 4HBS; gene; ds.
XX
OS Vanilla planifolia.
XX
FH Key Location/Qualifiers
FT CDS 1..1071
FT /tag= a
FT /product= "4HBS"
FT /transl_except= (pos:481..489,aa:Ser-Trp)
FT /transl_except= (pos:538..546,aa:Gly-Ser)
FT /transl_except= (pos:547..555,aa:Thr-Leu)
FT /transl_except= (pos:604..612,aa:Cys-Gly)
FT /transl_except= (pos:694..696,aa:Met)
XX
XX US2003070188-A1.
XX
XX 10-APR-2003.
XX
XX 28-FEB-2002; 2002US-00087714.
XX
XX 15-JUL-1997; 97US-0052604P.
XX 15-JUL-1996; 98WO-US014895.
XX 25-MAY-2000; 2000US-00462576.
XX 28-FEB-2001; 2001US-0272415P.
XX
XX (HVK/) HAVKIN-FRENKEL D.
XX (PODS/) PODSTOLSKI A.
XX (DIXO/) DIXON R A.
XX
XX Havkin-Frenkel D, Podstolski A, Dixon RA;
XX WPI; 2003-743794/70.
XX P-PsDB; ADC61377.
XX
XX Improving vanillin production in cultured Vanilla planifolia by

Db 700 GGCATTGAGTACATCAATAATGAGGAGTGCATACCGAGGAGTCTTACCCTTACAA 759
QY 687 TGGTGTCAATGTATCTGCAACTCAAGCAGAGAAATGTTGGTGTCAAGGTCATTTGATTC 746
Db 760 CGGTGTCAATGGCGTCTGCCATTACAAGGCTGAAATCCGTAGTTCAGGTTTGGACTC 819
QY 747 GATAAACATCACCTCGGTGGTGGTGGATGAGTTGAAGCATGCGAGTGGCTGGTGGCTGCC 806
Db 820 TGTTAACATCACTGAATGCTGAGACGAACCTGAAACGCCGCTGGTGGTTCGCCCC 879
QY 807 AGTTAGCGTTGCAATTTGAGTTGTCAAGGTTTCAATCTGTACAAAGAGGTGTATACAG 866
Db 880 AGTGAAGTGTTCCTTTAGTGTATCAACGTTTTCAGGCGAGTACAAGAGCGGAGTTTACAC 939
QY 867 CAGTGACACCTGTGAGAGATGCCAATGATGTGAACCAAGCAGTCTTCCGCTCGGTTA 926
Db 940 TAGCGACCACTGGGTACTACACCTGATGACGTGAACCAAGCAGTCTTCCGCTCGGTTA 999
QY 927 TGGAGTCGAGGACGGGATTCCTTATTGGCTCATCAAGAACTCATGGGTACAAATTTGGGG 986
Db 1000 CGGGTGGAGAACGGCGTCCCGTATTGGCTCATCAAGAACTCATGGGTACAAATTTGGGG 1059
QY 987 TGACAATGCTACTTTAAGATGGAACCTGGCAAGACATGTGTGTTGCAACTTGCCTGC 1046
Db 1060 TGACAATGCTACTTTAAGATGGAACCTGGCAAGACATGTGTGTTGCAACTTGCCTGC 1119
QY 1047 ATCTTATCCCATTTGGCTG 1066
Db 1120 GTCTTACCCCATCGTAGCGG 1139

RESULT 6
AAC37877
ID AAC37877 standard; DNA; 1358 BP.
XX
AC AAC37877;
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 18989.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131149P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132486P.

PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139452P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 21-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 21-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140931P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.

Db 901 GCATAGCATTTGAGGTTATACACTCGTTCGCGTTTACAGAGTGGAGTTTACACTGATA 960
Qy 872 ACACCTGTGGAAGAGATCCAATGGATGTGAACACCGAGTTCTTCCGCGTTCGGTTATGGAG 931
Db 961 GTCACGTGTGGAAGTACTCCAATGGATGTGAACACCGCGTTTTCGCGGTTGGTTATGGAG 1020
Qy 932 TCGAGGACGGGATTCCTTATTGGCTCATCAAGAACTCATGGGTACAAATTTGGGGTGACA 991
Db 1021 TTGAAGACGGTGTACCATATTGGCTTATTGAAGAACTCATGGGAGCGGATTTGGGGCGACA 1080
Qy 992 ATGGCTACTTTAAGATGGAACCTCGGCAAGAACATGTGTGGTGTGCAACTTGGCGCATCTT 1051
Db 1081 HAGTTACTTCAAGATGAGATGGGAGAGACATGTGTGGTATTGCTACATGTGCAATCAT 1140
Qy 1052 ATCCCATTTGGCT 1065
Db 1141 ACCCGTTGGCT 1154

RESULT 7
AAC48945
ID AAC48945 standard; DNA; 1370 BP.
XX
AC AAC48945;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 59352.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136292P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 22-JUL-1999; 99US-0145193P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 23-JUL-1999; 99US-0145276P.
PR 26-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.

1080 AAGGTTACTTCAAGATGGAGATGGGGAAGAACATGTGTGGTATTGCTACATGTGCATCAT 1139

XX 21-NOV-1996 (first entry)
 DT Up-regulated senescence clone, SENU3.
 XX
 DE Senescence related gene; expression; inhibition; acceleration; tomato;
 DE lettuce; cabbage; banana; strawberry; wheat; maize; rice; rape; soybean;
 KW delayed senescence; yield; protein content; quality; tolerance;
 KW increased senescence; desiccant; ss.
 XX Lycopodium esculentum.
 OS
 XX W09507993-A1.
 PN
 XX 23-MAR-1995.
 PD
 XX 13-SEP-1994; 94WO-GB001990.
 PF
 XX 13-SEP-1993; 93GB-00018927.
 PR
 XX (ZENE) ZENECA LTD.
 PA
 XX Smart CW, Thomas H, Hosken SE, Schuch WW, Drake CR, Grierson D;
 PI Parrell A, John I;
 PI WPI; 1995-131361/17.
 DR
 XX DNA constructs which modify expression of senescence-related genes -
 PT useful to accelerate or inhibit senescence in crop plants.
 PT
 XX Claim 7; Page 47-48; 78pp; English.
 PS
 XX The sequences given in AAT35095-133 are senescence related genes and
 CC fragments which were used in the construct of the invention which
 CC modifies the expression of at least one senescence related gene. Using
 CC these constructs senescence may be inhibited or accelerated in plants
 CC including tomato, lettuce, cabbage, banana, strawberry, wheat, maize,
 CC rice, rape or soybean. Delayed senescence may indirectly prolong the life
 CC of the plant, increase yield, increase protein content of fruits, improve
 CC quality of leafy vegetables, improve tolerance to disease, drought or
 CC other stress. Increased senescence may more rapidly break down unwanted
 CC plant material and so avoid the use of desiccants on crops. This sequence
 CC represents the full length sequence of the senescence up-regulating
 CC clone, pSENU3, which is also known as 77S3. pSENU3 is a cDNA of 1.1982 kb
 CC encoding a mRNA of approx. 1.4 kb. The mRNA encoded by pSENU3 is
 CC expressed during the onset of senescence in tomato leaves. pSENU3
 CC exhibits 70% homology with oryzain gamma, a cysteine protease expressed
 CC in rice seeds and induced by gibberellin, GA3. pSENU3 is deposited as
 CC NCIMB 40573
 XX
 XX Sequence 1388 BP; 394 A; 268 C; 329 G; 397 T; 0 U; 0 Other;
 SQ
 Query Match 44.4%; Score 475.8; DB 2; Length 1388;
 Best Local Similarity 68.5%; Pred. No. 4.8e-133;
 Matches 657; Conservative 0; Mismatches 302; Indels 0; Gaps 0;
 QY 106 CTTGATCGATGAGCCTGCCATCTCTCGCGCTCTTGGCAGTTGCGCCAGCCCTTCCAC 165
 DB 125 CCGATGAGTGGAGAACGGGATCTTCAAGTCTGCGGCCAGATCGCAGTCTCTCTCC 184
 QY 166 TTGCGACGTTGCGCCGCGAGTACGGAGAGCTACGGATCGGAGGAGATCAAGAAG 225
 DB 185 TTGCTGCTTTGCTATAGGCTATCGGAAAGGTATGATCTCGTTGAGAGATCAAGCA 244
 QY 226 AGGTTCCGGATCTTCGTGGAGATCTAGCGTTTATCCGGTCCACTAATCGGAGGATCTG 285
 DB 245 AGGTTTGAGATATTTTGGACATCTGAAGATGATCCGATCGCATCAAGAAAGGACTA 304
 QY 286 TCGTATACCTAGGATCAACCAATTCGCGACCTGACCTGGAGGAATTCGGGACCAT 345
 DB 305 TCATACAACTCGGTATCAATGAGTTTACCGACCTAACATGGATGAGTTCCGTAAACAC 364
 QY 346 CGCCTTGGTGGCGGAGAACTCTCGCGGAGTCTCGCGAGTCTGAGCAACCCGGTTTTCGAT 405

365 AAGTTGGGGGCATCTCAAAACCTGTTCTGCCACCAAAAGGGAAATCTCAAGTCTACTAAC 424
 QY 406 GCGGTGCTTCTCTGTACAGAGGATTTGGAGGGAGCAAGGATAGTGAGCCCTGTAAAGGAC 465
 DB 425 GTGTTCTGCCAGAGAGCAAGAGACTGGAGAGAGATGGTATTGTTAGCCAGTGAAGGCA 484
 QY 466 CAAGGAAGCTGGATCTTCTGCTGGACTTTTCAGTACTACTGGAGCACTAGAGCTGCATAT 525
 DB 485 CAGGGCAAGTGGGATCTTCTGCTGGACATTCAGCACTACTGGTGCACTAGAGGACGCTAT 544
 QY 526 ACACAGCTAACTGGAAAGAGCACATCATTTATCTGAACAGCAAACTTTGTGAGCTGTGCTCA 585
 DB 545 GCCCAAGCATTTGGGAAGGGAATCTCTCTGTGAGAGCAGCAGCTTGTGACTGTGCTGA 604
 QY 586 GCATTCAATAAATTGGATGCAATGAGGTTTGGCTTCCCAAGCCCTTTGAAATACGTTAAG 645
 DB 605 GCTTTTAAATAACTTTGGTTGCAATGAGGGTTGCCCTTCTCAAGCACTTGAAGTACATAA 664
 QY 646 TACAATGGAGGCATCGACACAGAAACAGACTTATCCATACCTTGGTGTCAATGGTATCTGC 705
 DB 665 TTCAATGGTGGTCTTGTATATTGGTGTAGCATATCCATACACCGGCAAGATGGCATATGT 724
 QY 706 AACTTCAAGCAGAGAAATGTTGGTGTCAAGGTCAATTTGATTCGATTAACATCACCCCTGGT 765
 DB 725 AAATTCTCACAGCAAAATATTGGTGTCAAAAGTCAATCAGTCTGTCAATATTACCCCTGGT 784
 QY 766 GCTGAGGATGAGTTGAAGCATGCAGTGGCTTGGTGTGCTCCAGTTAGCGTTGCAATTTGAG 825
 DB 785 GCTGAATATGAACTGAAATACGCACTGATGTTGGTGTAGGCTGTGCTGTGCTTTTGGAG 844
 QY 826 GTTGTGAAAGGTTTCAATCTGTACAAAGAGGTGTATACAGCAGTGCACCTGTGGAAGA 895
 DB 845 GTGGTAAAGGGTTCAACAGTACAAAGAGCGGAGTTTACGCCAGCACTGAAATGTGGCGAC 904
 QY 886 GATCCATGATGTGAACCAAGCAGTCTTCCGCTCGGTTATGAGTTCGAGTTCGAGGACGGATT 945
 DB 905 ACTCCCATGGACGTAAACCATGCTGTCTGTGCGGTTACGTTTGAATAATGTTACT 964
 QY 946 CCTTATTGGTCTCATCAAGAACTCATGGGGTACAAATTTGGGGTGCACAAATGGCTACTTTAAG 1005
 DB 965 CCCTACTGGCTATAAAGAACTCATGGGAGCAGATTTGGGGTGGAGATGGATACTTTCAA 1024
 QY 1006 ATGGAATCGGCAAGACATGTGTGTGTGCACTTGGCACTTGGCATCTTATCCCATTTGGC 1064
 DB 1025 ATGGAGATGGGAAAGAAATATGTGTGTGTGCAACTTGTGTGCACTTGTGCACTTGTG 1083
 RESULT 10
 AAC51435
 ID AAC51435 standard; DNA; 1077 BP.
 XX
 AC AAC51435;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 68510.
 XX
 KW Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway; metabolic pathway;
 KW promoter; termination sequence; ss.
 XX Arabidopsis thaliana.
 OS
 XX EP1033405-A2.
 PN
 XX 06-SEP-2000.
 PD
 XX 25-FEB-2000; 2000EP-00301439.
 PF
 XX 25-FEB-1999; 99US-0121825P.
 PR 05-MAR-1999; 99US-0123180P.
 PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 28-JUL-1999; 99US-0146386P.
PR 28-JUL-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.

XX 25-FEB-2000; 2000EP-00301439.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
XX 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145313P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 08-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149302P.
PR 23-AUG-1999; 99US-0149303P.
PR 23-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.

PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135623P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137503P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 22-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 08-AUG-1999; 99US-0147493P.
PR 08-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151067P.
PR 30-AUG-1999; 99US-0151130P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159584P.
PR 18-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161484P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.

PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
Query Match 40.1%; Score 429.4; DB 3; Length 2149;
Best Local Similarity 71.1%; Pred. No. 6.7e-119;
Matches 582; Conservative 0; Mismatches 236; Indels 1; Gaps 1;
QY 247 AATCTAGCGTTATCCCGTCCACTAATCGGAAGATCTGTGTATACCTAGGAAATCAAC 306
DB 1113 AATCTTGATTGATCAGATCCACCAAGAAAGCGTTATCTTACAACTCGGTGTTAAT 1172
QY 307 CAATTCGCGGACCTGACCTCGGAGGAATTCGGACCAATCGCTTGTGGCGGCAGAAC 366
DB 1173 CAATTTGCTGATTGACATGCAAGAGTTTCAAGGACCAAGCTTGTGCTGTCTCAGAAC 1232
QY 367 TGCTCGGCACTGCGCATGGAACACACCGGTTTGTGATGCGCTTCTCTGTAACGAGG 426
DB 1233 TGCTCTGCCACTTTAAAGGGCAGCCACCAAGGTCACAGAGCAGCTCTTCTCTGAAACAAA 1292
QY 427 GATTGGAGGAGCAAGGATAGTGAAGCCCTGTAAGGACCAAGAGCTGTGGATCTTGC 486
DB 1293 GACTGGAGAGAGATTGTATGTTAGTCCGTCAAGATCAGGAGGTTGTGATCTTGC 1352
QY 487 TGGACTTTCAGTACTTGGAGCACTAGAGCTGATATACACAGCTAACTGGAAGAGC 546
DB 1353 TGGACATTCAGCAGACTGGAGCTTTCAGGAGCTTACCATCAGGCAATTTGGAAAGGA 1412
QY 547 ACATCATTTCTGACACCACTTGTGACCTGCTCAGCATTAATACTTTGGATGC 606
DB 1413 ATAICTCTCTGAGCAACAGCTTGTGATGTTGCTGGAGCTTCAATTAATCTATGGTGC 1472
QY 607 AATGAGGTTTGCCTTCCCAAGCCCTTGAATACGTTAAGTAACTAGGAGGCAATCGACAC 666
DB 1473 AATGTTGCTTCTTCTCAAGCCCTTGAATACATAAATCAACGCTGGCTCGACACA 1532
QY 667 GAACAGACTTATCCATCTTGGTGTCAATGTTATCTGCACTTCAAGCAGGAGATGTT 726
DB 1533 GAGAAAGCTTATCTTATACCGGTAAGATGAAACCTGCAAAATTTTCAGCTGAAAGCGTT 1592
QY 727 GGTGTCAAGGTCATTGATTCGATAAATCAATCACTCCCTGGTGTGAGGATGATTTAAGCAT 786
DB 1593 GGTGTCAAGTCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 1652
QY 787 GCAGTGGCTTGGTGGTCCAGTTAGCTTGGATTTGAGTTGTGAAGTTTCAATCTG 846
DB 1653 GCGTTTGGATTGGTACGGCCAGTAAAGCATAGCATGCACTTGGAGTTATACACTGTTCCGCTT 1712
QY 847 TACAAGAAAGTTGATATACAGCAGTCAACCTGTGGAAGAGATCCAATGGATGTCAACAC 906
DB 1713 TACAAGAGTGGATTTACCTGATAGTCACTGTGGAAGTACTCCANTGGATGTGAACAC 1772
QY 907 GCAGTTCTTGGCTCGGTATGAGTGCAGGACCGGATTCCTTATTTGGCTCATCAAGAAC 966
DB 1773 GCGTTTGGCTGGTTGTTGATGGAGTTGAAGCAGGTTGACCATATTTGGCTTATTAAGAAC 1832
QY 967 TCATCGGTTCAAAATTTGGGTTGACAACTGCTTACCTTGAATGGAATCGGCAAGACATG 1026
DB 1833 TCATCGGAGCGGATTTGGGGGACAAAGGTTACTTCAAGATGGAGAT-GGGAAGACATG 1891
QY 1027 TGTGTTGTTCAACTTTGGGCTATCTTATCCCATTTGGCT 1065
DB 1892 TGTGTTGTTGCTATGTCATCATATACCCCGTTTGGCT 1930

RESULT 14

ADB94877

ID ADB94877 standard; cDNA; 1652 BP.

XX

AC ADB94877;

XX

DT 04-DEC-2003 (first entry)
XX Programmed cell death pathway cysteine protease cDNA #42.
DE
XX
KW programmed cell death; plant development; plant cell cycle; ATL2; DAD1;
KW Dnaase; lls; lsd1; nucellin-like aspartic protease; annexin; prohibitin;
KW fen-like protein; rac2; retinoblastoma-related protein; SINA; TFIID;
KW TEGT; xylogenic RNase; pur-alpha; cyteine protease; RPP5-like protein;
KW GP 91 NADPH oxidase subunit; NPR-like protein; BAG-1;
KW defender against cell death; lethal leaf spot; lesion stimulating death;
KW seven in absentia; transcription initiation factor;
KW testis enhanced gene transcript; gene; ss.
XX
OS Pinus radiata.
XX
PN US2003082724-A1.
XX
PD 01-MAY-2003.
XX
PF 14-AUG-2002; 2002US-00219220.
XX
PR 04-JUN-1999; 99US-00325932.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Flinn B, Lasham A;
XX
DR WPI; 2003-786916/74.
DR P-PSDB; ADB94919.
XX
PT New isolated polynucleotide useful for modulating programmed cell death,
PT altering the development cycle of plant cells, and subsequently modifying
PT plant development.
XX
PS Claim 1; Page 154; 214pp; English.
XX
CC The invention describes an isolated polynucleotide (I) comprising a
CC nucleotide sequence that is one of 145 fully defined sequences of 221-
CC 3415 base pairs (bp), given in the specification, and/or its complements,
CC reverse complements, reverse sequences, or sequences having 75, 90 or 95
CC % sequence identity to or that hybridise under stringent hybridisation
CC conditions to one of the 145 sequences. The methods and compositions of
CC the present invention to do with (I) are useful for modulating programmed
CC cell death and thereby altering the development cycle of plant cells, and
CC altering plant development. This sequence encodes a protein associated
CC with the programmed cell death pathway.
XX
SQ Sequence 1652 BP; 476 A; 318 C; 393 G; 465 T; 0 U; 0 Other;
Query Match 38.0%; Score 407.4; DB 9; Length 1652;
Best Local Similarity 65.7%; Pred. No. 2.7e-112;
Matches 626; Conservative 0; Mismatches 321; Indels 6; Gaps 2;
QY 116 TTGAGCTGCCATCTCTCGGCTCTTGGCAGTTGCGCCACGCTTCCACTTCGACCGT 175
DB 385 TGGAGTCCCAATTTCTGTTAAATCTTGGGAACCAATCCCAAGTCTTACAAATTCGGAGT 444
QY 176 TCGCCCGCAGGTACCGGAAAGAGCTACGATCGGAGGAGGAGATCAAGAGAGGTTCCGGA 235
DB 445 TCCTGTGAGATATGCAAGAGGTACGATTCGTCCATCAGCTTGTGCATAGATTCAATG 504
QY 236 TCTTCGTGAGATCTAGGTTTATCCGTTCACTTAATCGGAAGGATCTGTGATATACC 295
DB 505 TCTTTGTGAAGAACGTTGGAGCTGATCGAGTCAAGAAACAGAAATGAAGTTCCTTACTT 564
QY 296 TAGGAATCAACCAATTCGCGCACTTCCCTGGAGGAATTCGCGACCAATTCGCTTGGTG 355
DB 565 TGCATATATGATTTGCTGATACATGAGGAGGATTCATGACAAATATTTGGTG 624
QY 356 CGGCGCAGAACTGCTCGGAGCATGCGCATGGAACCAACCGGTTTGTGATGGGTCTTC 415
DB 625 CTTTACAGAACTGTTCCG---CTACCCACAGTAACCAATAAGTTGACGTATGCCAGCTTC 681


```

Db      742 CTGGCAAGGTCGCTGTCAAAAGTTGTGACTCTCTCAACATCTCTATGGGTGCTGAGGATG 801
Qy      776 AGTTGAAGCATGCAGTGGGCTTGGTGGTCCAGTTAGCGTTGCATTGAGGTTGTGAAAG 835
Db      802 AACTTAAAGCATGCAGTTGGCTGGTCCGGCCAGTCAGTGTGGCATTCCAGGTCACGGATG 861
Qy      836 GTTTCATCTGTACAGAAAGGTGTAT 862
Db      862 GCTTCCAGCTCTACGAGTCGGGTGTGT 888

```

Search completed: March 29, 2004, 19:38:48
Job time : 378 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 19:19:12 ; Search time 78 Seconds
(without alignments)
7619.903 Million cell updates/sec

Title: US-10-087-714-1

Perfect score: 1071

Sequence: 1 atggcagctaaagctctctt.....atccattgtggtgtgtag 1071

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfilest.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	541.4	50.6	1407	4	US-09-325-932A-107
2	280.2	26.2	677	4	US-09-325-932A-111
3	205.2	19.2	475	4	US-09-325-932A-120
4	161.4	15.1	1600	4	US-09-325-932A-98
5	160.2	15.0	1474	3	US-08-821-994-64
6	159	14.8	1833	3	US-09-120-365-6
7	159	14.8	1833	3	US-09-515-039-6
8	157.6	14.7	1056	1	US-09-500-651-1
9	157.6	14.7	1056	1	US-08-813-591-1
10	155	14.5	522	4	US-09-325-932A-112
11	150.6	14.1	1366	3	US-08-883-526-2
12	150.6	14.1	1366	4	US-09-519-283-2
13	147.4	13.8	1390	3	US-08-821-994-61
14	147.4	13.8	1441	3	US-08-821-994-63
15	145.8	13.6	1102	3	US-08-821-994-86
16	144.2	13.5	1434	3	US-08-821-994-62
17	139.4	13.0	990	2	US-08-806-959-1
18	137.2	12.8	1661	3	US-08-821-994-82
19	134	12.5	1575	4	US-09-784-642-1
20	131.6	12.3	1553	3	US-08-821-994-60
21	127.6	11.9	1463	4	US-09-325-932A-92
22	125	11.7	1482	1	US-08-330-1218-1
23	125	11.7	1482	1	US-08-536-861-1
24	125	11.7	1482	5	PCT-US95-13820-1
25	125	11.7	1614	4	US-08-684-932A-35
26	125	11.7	1619	1	US-08-208-007A-1
27	125	11.7	1619	4	US-08-915-095A-1

28	125	11.7	1619	4	US-08-798-096-1	Sequence 1, Appl
29	125	11.7	1619	4	US-08-798-096-1	Sequence 1, Appl
30	125	11.7	1619	4	US-09-953-956-1	Sequence 1, Appl
31	125	11.7	1619	4	US-08-553-125A-1	Sequence 1, Appl
32	125	11.7	1619	4	US-10-114-464-1	Sequence 1, Appl
33	125	11.7	1619	5	PCT-US94-04781-1	Sequence 1, Appl
34	125	11.7	1669	3	US-08-964-308-3	Sequence 3, Appl
35	125	11.7	1669	3	US-08-964-313-3	Sequence 3, Appl
36	125	11.7	1669	4	US-09-069-138-3	Sequence 3, Appl
37	123.4	11.5	990	3	US-08-964-308-12	Sequence 12, Appl
38	123.4	11.5	990	3	US-08-964-313-12	Sequence 12, Appl
39	123.4	11.5	990	4	US-09-069-138-12	Sequence 12, Appl
40	121.8	11.4	990	3	US-08-964-308-13	Sequence 13, Appl
41	121.8	11.4	990	3	US-08-964-313-13	Sequence 13, Appl
42	121.8	11.4	990	4	US-09-069-138-13	Sequence 13, Appl
43	121.8	11.4	1577	3	US-08-821-994-59	Sequence 59, Appl
44	118.6	11.1	660	3	US-09-004-731-94	Sequence 94, Appl
45	118.6	11.1	660	3	US-08-749-699-94	Sequence 94, Appl

ALIGNMENTS

RESULT 1
US-09-325-932A-107
; Sequence 107, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 107
; LENGTH: 1407
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-09-325-932A-107

Query Match	50.6%	Score	541.4	DB	4	Length	1407
Best Local Similarity	70.5%	Pred. No.	1.8e-157	Indels	3	Gaps	1
Matches	738	Conservative	0	Mismatches	306		
Qy	21	CTTCCTACTCTTCTCTGCTCTCGCCCTCTCCGTCGCGCTCGCGGTTTCGAGAGACAA	80				
Db	87	CGTCCTCTCTCTGCTCTCGCGCTCTCTCTCGCGCGCGGCTCGAGCTTCGAGAGTCCAA	146				
Qy	81	TCCAAATCCGTCGTTTACACA---AAGGCTGACTGATGAGCTGCCATCTCTGGCGT	137				
Db	147	CCCATCTCCGCTCTTCCCGAGCGGCGCTCCGCGACCTCGAGTCTCCATGCTCAGAT	206				
Qy	138	CCTTGCAGTTGCGCGCAGCCCTTCCACTTCGACAGGTTTCCGCGCAGTACGGGAAGAG	197				
Db	207	CGTCGCGCGCAGCCCGCCAGCCCTTCTCTCTCGCGCTTCGCCAACAGGTATGGGAAGAG	266				
Qy	198	CTACGAGTCGAGGAGAGATCAAGAGAGTTTCGGATCTTCGTCGAGAACTACGGTT	257				
Db	267	GTACGAGACCGCGGAGGAGATCAAGTCTCGGTTTCGAGATCTTCAGGAGAACTCAAGTT	326				
Qy	258	TATCCGCTCCACTAATCGGAAGGATCTGTCGTATACCTTAGGAATCAACCAATTCGCCGA	317				
Db	327	GATCCGATCCACCAACAAGAGGGCTTGCCCTACACCTCGGTGTCAATAAGTTCTGTA	386				
Qy	318	CTGAGCTGGAGAGAAATTCGGACCAATCGCTTCGTCGCGCGCAGAACTGTCGGCGAC	377				
Db	387	TTGGAGCTGGAGAGATTACAGGAGGACAGACTGGGAGTGTCTCAAACTGTCTGCCAC	446				
Qy	378	TGCGCATGGAACACCGGTTTCTCGATGCGCTTCTCTGTAAACAGGAGATTGGAGGA	437				

```
Db 447 CACCAAGGGCAACCAACAGCTCACCGACGAAGCTCTTCCCGAGATGAAGAAGCTGGAGAGA 506
Qy 438 GCAAGGGATAGTAGAGCCCTGTAAAGGACCAAGGAAGCTGTGGATCTTGTGACTTTTCAG 497
Db 507 AAAGGGCATGTAAAGCCAAATTAAGATCAGGGGCACTGTGGATCTTGTGACTTTTCAG 566
Qy 498 TACTACTGAGCACTAGAGGCTGCGATATACAGAGCTAATCTGGAAGAGACACATCATATC 557
Db 567 TACCACCTGAGCTCTTGAAGGCTGCTTATCACAAGCATTTCGGGAACAAATCTCTCTGTC 626
Qy 558 TGAACAGCAACTGTGGACTGTGCTCAGCATTTCAATTAATTTGATGCAATGGAGGTTT 617
Db 627 TGAGCAGAGCTGCTGAGATCTGTGCGGCTTTCAACAACTTTGATGTAGTGGTGGACT 686
Qy 618 GCCTTCCCAAGCTTTGAATACGTTAAATGAGGAGCATCGACACAGAAACAGACTTA 677
Db 687 GCCATCCCAAGCTTTGAGTACGTCAAGTACAAACGGTGGCTTGTATCCGAGGAAGCATA 746
Qy 678 TCCATACCTTGGTCAATGGTATCTGCAACTTCAAGCAGGAGAAATGTTGGTGTCAAGT 737
Db 747 TCCTTATACCGCAGTGGATGTAGCTGCAATTTCTCGGCTGATATGTTGGTGTCAAGT 806
Qy 738 CATTTGATTCGATAAATCACCTCGGCTGTGAGGATGAGTTGAAGCATGCAAGTGGCTT 797
Db 807 GCTCGACTCTGTTAAACATCACCTTGGTGTGAGGATGAATTAAGCATGCAAGTGGCTT 866
Qy 798 GGTGGCTCGATTTAGGTTGCACTTGAAGTGTGGAAGGTTTCAATCTGTACAGAAAGG 857
Db 867 CGTCCGGCCAGTGAAGTGTGCACTTCCAGGTCGTGGAAGACTTCAGATTTACAGTGGG 926
Qy 858 TGTATACAGCAGTGACACCTGTGGAAGAGATCCAAATGGATGGAACCAAGCAGTCTTTCG 917
Db 927 TGCTACACAGCGATACATGCGTAGCATTCCATGGATGGAACCATGCTTCTGCG 986
Qy 918 CGTCGGTTATGAGTGTGAGACCGGATTCCTTATTTGCTCATCAAGAACTCATGGGTAC 977
Db 987 TGTGTTATGAGGTTGAAGATGGTGTTCGTTCTCGCTCATCAAGAAATTCCTGGGAGC 1046
Qy 978 AAATTTGGGTGACAAATGGCTTAAAGATGGAACCTCGGCAAGAAACATGTGTGGTTCG 1037
Db 1047 AGACTGGGTGACCAAGGATCTTCAAGATGGAGTGGGAAGAAACATGTGTGGATCGC 1106
Qy 1038 AACTTGGCATCTTATCCCAATTTGGC 1064
Db 1107 TACTTGGCATCATACCTGTTGGC 1133

RESULT 2
US-09-325-932A-111
; Sequence 111, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: death and their use in the modification of forestry plant develo
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 111
; LENGTH: 677
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-09-325-932A-111

Query Match 26.2%; Score 280.2; DB 4; Length 677;
Best Local Similarity 68.1%; Pred. No. 9.6e-77;
Matches 390; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

Qy 116 TTGAGCTCGCATCCTCGGCGCTCTTGGCAGTTGCGCCACGCTTCCACTTCGACGGT 175
```

```
Db 102 TCAGTCTCTCCATCTCTCCAAACCGTTGGCCACGCGCGCTCCCGCCCTCTCTCTTCTAGACT 161
Qy 176 TCCCCGCAGGTACGGGAAGAGCTACGGATCGGAGGAGAGATCAAGAAAGAGGTTTCGGGA 235
Db 162 TTCCAGCAGGTACGAGAAGAGGTACGAGACAGCGGAGGAGATCAAGTTTGAGGTTTCGATA 221
Qy 236 TCTTCGTGGAGAACTTAGCGTTTATCCGFTCCACTAATCGGAAGGATCTGTTCGTATACCC 295
Db 222 ATTACAGGAGAACTTCAAGCTCATTTGATCCACCAACAGAGGGCTTGCCTTACACTC 281
Qy 296 TAGGAATCAACCAATTTGCCGACCTGACCTGGGAGGAAATTCGGACCAATCGCTTTGGTG 355
Db 282 TCGCTGTTAATCAGTATGCTGACTGGAGCTGGGAGGAGTTCAAGACGCAACAGACTGGGAG 341
Qy 356 CGGCGCAAGACTGCTCGGCACTGCGCATGGAACCAACCGGTTTGTTCGATGGGCTGCTTC 415
Db 342 CTTCTCAAGACTGCTCTGCCACCAACAGGGGAGCCCAAGCTCTCGAGGCTGCTTATCACCAAGCAC 401
Qy 416 CTGTAACAGGAGGATGAGGAGCAAGGATAGTGAAGCCCTGTAAGGACCAAGGAAGCT 475
Db 402 CCAAAACGAAAGACTGGAGAAAGAGGGCAATGTAAGCCCAAGTTAAAAATCAAGGGCGCT 461
Qy 476 GTGATCTTCTGAGACTTTTCACTACTGAGCACTAGAGGCTGACATATACACAGCTAA 535
Db 462 GTGATCTTCTGAGGTTTTCAGGCGCACTGAGGCTCTCGAGGCTGCTTATCACCAAGCAC 521
Qy 536 CTGAAAGAGACACATCATTTATCTGAACAGCAACTTTGTGACTGTGCTTCAGCAATTCATA 595
Db 522 ACGGGAAGGAATCTCTCTGCTGAGCAGCAGCTCGTGGACTGCGCTACGCTTTCAACA 581
Qy 596 ACTTTGATGATGAGGAGTTTTCCTTCCCAAGCTTTGATAGCTTTGAATAGTAACTAGAG 655
Db 582 ACTTTGATGATGAGGAGTTTTCCTTCCCAAGCTTTGAGTAACTAGTAACTAGAG 641
Qy 656 GCATCGACACAGAACAGACTTTATCCATACCTTG 688
Db 642 GCCTTGAGACCGGAAGCTTATCCTTATCTG 674

RESULT 3
US-09-325-932A-120
; Sequence 120, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: death and their use in the modification of forestry plant devel
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 120
; LENGTH: 475
; TYPE: DNA
; ORGANISM: Pinus radiata
US-09-325-932A-120

Query Match 19.2%; Score 205.2; DB 4; Length 475;
Best Local Similarity 66.0%; Pred. No. 1.3e-53;
Matches 313; Conservative 0; Mismatches 158; Indels 3; Gaps 1;

Qy 151 CGCCAGCGCTTCCACTTCGCACCGTTCGCCCGCAGGTACGGGAAGAGCTACGGATCGGAG 210
Db 5 CCCAAGTCTTAGAGTTTGGAGTTTGGTCTCAGATATGCAAGAGAGTACGATTTCTGTC 64
Qy 211 GAGGAGATCAAGAAGAGGTTTCGGATCTTCTGAGAAATCTAGCGTTTATCCGGTCCACT 270
Db 65 CATCAGCTTGTGCATAGATTCAATGCTCTTTGTGAAGAACGTGGAGCTGATCGAGTCAAGA 124
Qy 271 RATCGAAGGATCTGCTGATACCTTAGAATCAACCAATTCGCGCACTGACCTCGGAG 330
```

Db 125 AACAGATGAAGCTTCCTTATCTTGGCAATAAATGAGTTTGTGACATAAATCATGGAG 184
Qy 331 GAATTCGGACCAATCGCTTGGTGGCGGAGAACTGTCTGGGAGATGCGCATGGAAC 390
Db 185 GAATTCATGACAAATATTTGGGTGCTTCAAGAACTGTTTCGG---CTACCCACAGTAAC 241
Qy 391 CACCGGTTTTCGATGGCGTCTTCCCTGTAAACGAGGATTTGGAGGAGCAAGGATAGTG 450
Db 242 CATAGTTGAGTATGCCAGCTTCCTCGNAGAAAGCTGGAGACAAAGAGGCATAGTG 301
Qy 451 AGCCTGTAAAGGACCAAGGAGCTGTGGATCTTGTCTGGACTTTCACTACTCTGAGCA 510
Db 302 AGTCTGTAAAGGACCAAGGAGCTGTGGATCTTGTCTGGACTTTCACTACTCTGAGCA 361
Qy 511 CTAGAGCTGCATATACAGCTAACTGGAAGGACATCATATCTGACACGCACTT 570
Db 362 CTAGAGCTGCATATACAGCTAACTGGAAGGACATCATATCTGACACGCACTT 421
Qy 571 GTGAGCTGTGCTCAGCAATCAATTAATTTGGATGCAATGGAGTTTGCCTTCC 624
Db 422 GTTGAATGTGCTGAGCAATTAACAACTTTGGTTGCAATGGTGGACTGCCATCC 475

RESULT 4
US-09-325-932A-98
; Sequence 98, Application US/09325932A
; Patent No. 6451804
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Amette
; TITLE OF INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: death and their use in the modification of forestry plant develop
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 98
; LENGTH: 1600
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-09-325-932A-98

Query Match 15.1%; Score 161.4; DB 4; Length 1600;
Best Local Similarity 52.9%; Pred. No. 1e-39;
Matches 460; Conservative 0; Mismatches 376; Indels 33; Gaps 4;

Qy 165 CTTCGACGGTTCCGCCAGGTACGGAGAGCTAGCGATCGGAGGAGAGATCAAGAA 224
Db 244 CTAGAGAGCTGGCTCGCAAGACGGCAAGGCTTACACGCCCTGGCGGAGAGAGAA 303
Qy 225 GAGGTTGGGATCTTCGTGGAGATCTAGCGTTTATCCGGTCCACTAA---TCGGAAGGA 281
Db 304 GCGCTTCAGGTCTTCAAGGACAACTCCGGTTTCATCGACGACCAACGCCGCGGGGA 363
Qy 282 TCTGTGCTATACCTTAGGAATCAACCAATTCGCGAGCTGACCTGGGAGGAATTCGCGAC 341
Db 364 CGGACCTACAGGTGCGCTTCAACGATTCGCGAGCTTCACTAACGAGAGTACCGGTC 423
Qy 342 CAATGCGCTTGGTGGCGCGAGAACTGCTCGGAGCTGCGCATGGAACCAACCG- 395
Db 424 CATGTACCTGGCGCGAGGAGCTCGGTGGGCGCGGCTCGGAGGGGCGCGCAGCA 483
Qy 396 -----GTTTGTGATGGCTTCTTGTAAACGAGGATTTGGAGGGAGCAAGG 443
Db 484 TCGGTACCGCTGGCGCGGAGGAGCTCGCGGCTCGGTGATTTGGAGAGAGG 543
Qy 444 GATAGTACCGCTGTAAGGACCAAGGAGCTGTGGATCTTGTCTGGACTTTTCAGTACTAC 503
Db 544 CGCGCTTGTGACGTCAAGGACCAAGGAGCTGCGGAGTTGCTGGGCGTTCTCTCAAT 603
Qy 504 TGGAGCACTAGAGGCTGCATATACAGCTAACTTGGAAAGGACATCATATCTGAC 563

Db 604 TCCTGCTGGAGGGGATAAACAAGCTTGTGACTGGTGTGATCTCTCTGTCCGAGCA 663
Qy 564 GCAACTGTGTGAGCTGTGCTCAGCAATTCATAAATTTGGATGCAATGGAGGTTTGCCTTC 623
Db 664 GGAATCTGTGGAGTCCGATACATCTCATAATGA---AGGATGTAAATGGCGGCTCATGGA 720
Qy 624 CCAAGCCTTTGAATACGTTTAAGTACAAATGGAGGATCGACACAGACAGACTTATCCATA 683
Db 721 TTATGCCCTTTGAATTCATATCAACACGAGGCAATGATACCGAGGAGATATCCCTA 780
Qy 684 CTTGGTGTCAATGTATCTGCAATTCAGCAGGAGAAATGTTGGTGTCAAGGTCAATGA 743
Db 781 TAGAGCTGTAGATAGCACTTGTGACCAATACAGGAAGAACGCAAGGTTGTGACGATTA 840
Qy 744 TTCGATAAACAATCACTCCCTGGGTGCTGAGATGAGTTGAGCATGCAAGTGGGCTTGGTGG 803
Db 841 CGATTTATGAGATGTTCCAGAAATGATGAGAAAGCATTCGAAAGGCGAGTTGCTTAATCA 900
Qy 804 TCCAGTTAGCTGTGCAATTTGAGTTTGTAAAGTTTCAATCTGTACAGAAAGGTGTATA 863
Db 901 ACCAGTCACTGTGGCAATTTGAAGCAGGAGCGCGGAAATTCAGTTTATGATTCGGGTAT 960
Qy 864 CAGCAGTGACCTCTGGNAGAGATCCAATGATGTAACACGAGCTTCTTGGCGTGG 923
Db 961 ATTTACTGGCAATGTGGGACAGCTC-----TGGATCATGGGTTTACTGCGAGTCGG 1011
Qy 924 TTATGGAGTCGAGGACGGGATTCCTTATTTGGCTCATCAAGAACTCATGGGTGACAAATG 983
Db 1012 ATATGGCAGAGAAACGGAGTTGATTCTGGATAGTGAAGAACTCATGGGCGGTAGCTG 1071
Qy 984 GGGTGCAATGCTGCTACTTTAAGATGGAAC 1012
Db 1072 GGGAGAGCAAGGTTACATCAAAATGGCAC 1100

RESULT 5
US-08-821-994-64
; Sequence 64, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 64
; LENGTH: 1474
; TYPE: DNA
; ORGANISM: Brassica napus
US-08-821-994-64

Query Match 15.0%; Score 160.2; DB 3; Length 1474;
Best Local Similarity 56.5%; Pred. No. 2.3e-39;
Matches 344; Conservative 0; Mismatches 253; Indels 12; Gaps 2;

Qy 402 CGATTCGCTGCTTCTGTAAACGAGGATTTGGAGGAGCAAGGATAGTGAGCCCTGTATAA 461
Db 455 CGAGCTGGAGGTTCCGAGACCGTTGATTGGAGACGGAAGAGGCGGTTAATGCCATTAA 514
Qy 462 GGAACCAAGGAGCTGTGGATCTTGTGGACTTTTCACTACTCTGAGAGCTAGAGGCTGC 521
Db 515 AAACCAAGGAGCTTGGCGAAGTTTGTGGCGTTTTCGACAGCTGCGAGCAGTAGAAGGTAT 574
Qy 522 ATATACAGCTAACTTGGAAAGGACATCATATCTGAAACGAACTTGTGACTGTGC 581

575 AACAAAGATCGTAAACAGGAGAACTCATATCTCTGTCCGAAACAGAACTTGTTCGACTCGGA 634
582 CTCAGCATCAATACTTTGGATGCAATGGAAGTGTTCCTTCCCAAGCCTTTGAATACGT 641
635 C---AGATCCTCAACCAAGGCTGCAACGGTGGTTAAATGGAATATGCTTTTCAATTCAT 691
642 TAAGTACAAATGGAGGATCGACACAGAAACAGACTTATCCATACCTTGGTGTCAATGGTAT 701
692 CATGAATAACGGGGTTTGAACACCGAGCAAGATTATCCTTACCGTGGTTCATATGGA 751
702 CTCAACATTTCAAGCAGGAGAAATTTGGTGTCAAGGTCAATGATTCGATTAACATCACCCCT 761
752 ATSCAATCTTTACTGAAGATTTCAAGAGTGTGAACATTTGATGGTTACGAAGATGTTCC 811
762 GGTGTCTGAGGATGATGAGCATGAGTGGCTTGGTGGCTTCCAGTTAGGTTGCAAT 821
812 TACTGAAGATGAACCGGCTTGAAGAGAGCAGTTTCATACCGCCGCTGAGTGTGGCCAT 871
822 TGAGGTTGTGAAGGTTTCAATCTGTACAAAGAGGTGTATACAGCAGTGACACCTGTGG 881
872 TGAAGCTGGTGAAGAGTTTCCAACTTACCAATCGGGGATCTTCACTGGAAGAGTGG 931
882 AAGAGATCCAAATGGATGTGAACACGAGTTCTTGGCTGCGTTATGGAGTCGAGACGG 941
932 GACAAATC-----TAGATCATGCGTGGTGGCTGTGGTTATGTTTCAGAGAACGG 982
942 GATTCTTATTGGCTCATCAAGAACTCATGGGCTCAAAATTTGGGGTGACAAATGGCTACTT 1001
983 TATTGACTATTGATTGTAAAGAACTCGTGGGCTACAGTTGGGAGAGGTTGGTTACAT 1042
1002 TAAGATGGA 1010
1043 TAGGATGGA 1051

RESULT 6

US-09-120-365-6
; Sequence 6, Application US/09120365
; Patent No. 6103514
; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/120,365
; EARLIER FILING DATE: 1998-07-22
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1833
; TYPE: DNA
; ORGANISM: Periplaneta americana
US-09-120-365-6

Query Match 14.8%; Score 159; DB 3; Length 1833;
Best Local Similarity 51.0%; Pred. No. 6.1e-39;
Matches 462; Conservative 0; Mismatches 425; Indels 18; Gaps 3;
166 TTCACAGGTTGCGCCGAGTACGGAAGAGCTACGGATCGGAGGAGGAGATCAAGAAG 225
812 TTTGATCATTCGAAAAAGACACAGCAAGACTACGCCAGCAATTTGGAACACACAAAG 871
226 AGTTCCGGATCTTCGTGGAGATCTAGCGTTTATCCGGTCCACTAATCGAAGGATCTG 285
872 AGAAAAAGAAATTTTCGACAGAAATTTGAGGTTTATCCATTTAGAAATCGTGTAGATTA 931
286 TCGTATACCTTAGGAATCAACCAATTCGCCGACCTGACCTGGGAGGAATTC----- 337
932 GGATTCATCTGGAGCTGAACCAATTTGGCGACCGGACAGAGCTCGAACTGAAAGCTCTG 991
338 ----GGACCAATCGCTTGGTGGCGGAGAACTGCTCGGCGACTGGCATGGAACCCAC 393

992 AGGGGAAGCAATACATGATGGGTACATGAGGTTCTCCATTTCTCTTATACCAATCTT 1051
394 CGGTTTTCGATGGCGTGTCTCTGTAACGAGGATTTGGAGGAGCAAGGATAGTGAGC 453
1052 GAGCAATCATGGACCAAAATTCCTGATGATTTGGACTGGAGAAATTTATGTTGCTGACT 1111
454 CCTTAAAGACCAAGGAAGCTGTGGATCTTTGTGGACTTTTCACTACTCTGGAGCACTA 513
1112 CCAGTTAAAGATCAGTCTGTTTGGGCTCTTTGTTGAGTTTGGACTACTGGCACCATC 1171
514 GAGCTGCATATACACAGCTAACTGGAAGAGACATCATTTATCTGAACGAACTTTGTG 573
1172 GAAGGAGCTTATTTCTTAAAGATGAGCAATTTGGTGGATTTGTCAACACAGGCTCTAAT 1231
574 GACTGTGCTCAGCATTCATTAATTTTGGATGCAATTTGGAGGTTTGGCTTCCCAAGCCTT 633
1232 GACTGCACTGGGCTATGTAACATTTGTTGATGAGGTTGAAGATTTCCGCTCTTAT 1291
634 GAATACCTTAAAGTACAATGAGGATCGACACAGAAACAGACTTATCCA---TACCTTGGT 690
1292 GAATGATGATGAAGCATGGTGGCATCCCACTGGAAGACGAATATGAGGCTATTTGGGC 1351
691 GTCATGCTATCTCAACTTCAAGCAGGAGAAATTTGGTGTCAAGGTCAATGATCGATA 750
1352 CAGGATGCTATTGTCTGTAATAATGTAATCTCTACAGCAAGATCACTGGCTATGTG 1411
751 AACATCACTTGGTGTGCTGAGGATGAGTTGAAGCATGAGTGGCTTGGTGGCTCCAGTT 810
1412 AACGTCATCTGAGACATTTGATGCTTTGAAGTAGCACTAGCAAGCATGGTCCCATC 1471
811 AGCGTTGCATTTGAGTGTG---AAAGGTTTCAATCTGTACAGAAAGGTTGATACAGC 867
1472 TCTGTGCAATTTGATGCTTCTCAAGAACCTTCTCTTCTTATTCGAATGGCATCTACT 1531
868 AGTGACACCTGTGAAGAGATCCAAATGGATGTGAACACCGAGTTCTTGGCGTGGTAT 927
1532 GATCTGAGTGTGAAACAACTAGATCAGTTGACCAAGCAGTTCTGCTGGTTGGTAT 1591
928 GGAGTCGAGGAGGATCTCTTATTTGCTCATCAAGAACTCATGGGTACAAATGGGGT 987
1592 GGCATCAATAATGAAACCCCTTATTTGGTGGTCAAAATTTCTGCTCTAATTTATTTGGGC 1651
988 GACAAATGCTACTTTAAGATGGAATCGGCAAGCAATGTGTGGTGTTCGAATTTGGCA 1047
1652 AACGATGGCTACATCTTATGTCACCTAAGGATTAACAATGCTGTGTTCGAATGACCT 1711
1048 TCTTA 1052
1712 ACCTA 1716

RESULT 7

US-09-515-039-6
; Sequence 6, Application US/09515039
; Patent No. 6214599
; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/515,039
; CURRENT FILING DATE: 2000-03-06
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1833
; TYPE: DNA
; ORGANISM: Periplaneta americana
US-09-515-039-6

Query Match 14.8%; Score 159; DB 3; Length 1833;
Best Local Similarity 51.0%; Pred. No. 6.1e-39;

Matches 462; Conservative 0; Mismatches 425; Indels 18; Gaps 3;
QY 166 TTCCGAGGTTCCGCCGAGTACCGGAAGAGTACGATCGGAGGAGGATCAAGAG 225
Db 812 TTTGATCATTCCGAAAAGACACAGCAAGAGTACGCCAGCAATTTGGAAACACAAAG 871
QY 226 AGGTTCCGGATCTTCGTGAGAAATCTAGCGTTTATCCGTCCTCAATATCGGAAGAGTCTG 285
Db 872 AGAAAGAAATTTCCGACAGAAATTTGAGGTTTATCCATTTCTAAGAAATCGTGTAGATTA 931
QY 286 TCGTATACCTTAGAAATCAACCAATTCGCCGACCTGACCTGGGAGGAAATCC----- 337
Db 932 GGATTCATCTCGGACGTGAACCAATTTCCGCGACCGGACAGAGCTCGAACTGAAAGTCTG 991
QY 338 ----GGACCAATCCCTTGGTGGCGCGCAGAACTGCTCGCGACTGCGCATGGAACCAAC 393
Db 992 AGGGGGAAGCAATACATGATGGGTACATGGAGGTTCTCCATTTCTTATACCAATCTT 1051
QY 394 CGTTTTCGATCGGCTTCCTGTAAACGAGGATTTGGAGGAGCAAGGATAGTGAAC 453
Db 1052 GACGCAATCATGGACCAAAATTCCTGATGATTTGACTGGAGAAATTTATGGTGTGACT 1111
QY 454 CCTGTAAGGACCAAGGAGCTGGATCTTCTGGACTTTCAGTACTACTGGAGCACTA 513
Db 1112 CCAGTTAAGATCAGTCTGTTTGGCGCTCTTGTGGAGTTTGGGACTACTGGACCAATC 1171
QY 514 GAGGCTGCATATACACAGCTAACTGGAAGAGCACATCATTTCTGAAACAGCAACTTGTG 573
Db 1172 GAAGGAGCTTATTTCTTAAAGTATGGACATTTGGTGGCAATGTCACAAAGGCTCTAAT 1231
QY 574 GACTGTCCCTCAGCATTCATTAATCTTGGATGCAATGGAGGTTTCCCTCCCAAGCTTT 633
Db 1232 GACTGCACTGGGCTATGGTAAATGTTGTTGATGGAGTGAAGATTTCCGCTCTTAT 1291
QY 634 GAATACGTTAAGTAACTGAGGAGCATCGACACAGACAGACTTATCA---TACCTTGGT 690
Db 1292 GAATGGATGATGAAGCATGGTGGCATCCACTGGAAGACGAATATGGAGCTATTTGGGC 1351
QY 691 GTCAATGATATCTGCAACTTCAAGCAGAGAAATGTTGGTGTCAAGGTCAATGATCGATA 750
Db 1352 CAGGATGGCTATTGTCTATGTCGAAATGTAACTCTCACAGCAAAAGATCACTGGCTATGT 1411
QY 751 AACATCACTCCCTGGTGTGAGATGAGTTGAAGCATGAGTGGGCTTGGTCCGTCAGTT 810
Db 1412 AACGTCACTCTGAGACATTTGATGCTTTGAGGTAGCATAGCAAGCATGTGCCATC 1471
QY 811 AGCGTTGCAATTTAGGTTGTG---AAAGGTTTCAATCTGTACAGAAAGGTGTATACAGC 867
Db 1472 TCTGTGCAATGATGCTTCTCACAGACCTTCTCCTTCTATTCGAATGGCATCTACTAT 1531
QY 868 AGTCACTCTGGAAGAGATCCATGGATGTGAACCAAGCATCTTCCGTCGGTTAT 927
Db 1532 GATCCTGAGTGTGAAAACAGCTAGATCAGTTGACCAAGCATGCTGCTGTTGGTTAT 1591
QY 928 GGAGTTCGAGGACGGGATTCCTTATTTGGCTCATCAAGAACTCATGGGTACAAATTTGGGT 987
Db 1592 GGCATCAATAATGAAACCTTATTTGGTGTGCTCAAAATTCCTGTCTAATTTATTTGGGC 1651
QY 988 GACAAATGCTACTTTAAGATGGAATCTCGGCAAGCAATGTTGTGTGTGTGCACTTGGCA 1047
Db 1652 AACGATGGCTACATCTTATGCTACCTAAGGATAACCACTGTGTGTGTGCACTGACCCT 1711
QY 1048 TCTTA 1052
Db 1712 ACCTA 1716

RESULT 8

US-09-500-651-1
; Sequence 1, Application US/09500651
; GENERAL INFORMATION:
; APPLICANT: ASANO, MINAO
; APPLICANT: KAWAI, MISAOK

APPLICANT: MIWA, TETSUYA
APPLICANT: NIO, NORIKI
TITLE OF INVENTION: AMINOPEPTIDASE GX, AND A METHOD OF
TITLE OF INVENTION: HYDROLYSING A PROTEIN WITH THE SAME
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSES: OBOLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESSER: P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/500,651
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/813,591
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 030458/1997
FILING DATE: 14-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: OBOLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-845-0
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1056 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Glycine max
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1056
OTHER INFORMATION: /note= "METHOD OF DETERMINING THE
OTHER INFORMATION: CHARACTERISTICS: P"
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..9-
OTHER INFORMATION: /note= "INSERTION SEQUENCE, METHOD
OTHER INFORMATION: OF DETERMINING THE CHARACTERISTICS: E"
US-09-500-651-1

Query Match 14.7%; Score 157.6; DB 1; Length 1056;

Best Local Similarity 52.7%; Pred. No. 1.2e-38;

Matches 455; Conservative 0; Mismatches 379; Indels 30; Gaps 4;

QY 166 TTCGACAGGTTCCGCCGAGGTTACGGAAGAGTACGATCGGAGGAGGATCAAGAG 225

Db 73 TAGGACAGTGGCTCGTGAAGCACGGAAGGTGTCAACGCGCTCGCGCGAAGAGAG 132

QY 226 AGGTTCCGGATCTTCGTGGAGAAATCTAGCGTTTATCCGTTCCACTAA---TCGGAAGGAT 282

Db 133 CGCTTCCAGATCTTCAAGGACACACCTCGGATTCATCGACGACCAACATCCCGGAGGAC 192

QY 283 CTGTGCTATACCTTAGGAATCAACCAATTCGCGGACCTGACCTGGAGGAAATTCGGACC 342

Db 193 CGAAACCTACAAGCTCGGACTGAACCGTTTCGCTGATCTCACCACAGGAAATACAGGGCC 252

QY 343 AATCGCCTGTTGTGCGCGCAGAA-----CTGCTCGCGGACTGCGCATGGA 387

253 AAGTACTTTGGGAACCAAGATGATCCCAACCGGAGGCTCGAAGAGACCCCGAGCAACCGC 312
388 AACCAACCGGTTTGTGGATGGCGTCTCTCTGTAAACGAGGATTTGGAGGAGCAAGGGATA 447
313 TAGCGCCACGCTGTGGGACAAATTTGCTGATTCGGTTGATTTGGAGGAAGGAGTGCT 372
448 GTGAGCCCTGTAAAGGACCAAGGAGCTGTGGATCTTCTGCTGAGCTTTCAGTACTACTGGA 507
373 GTTCTCTCTGTCAAGACCAAGGAGGCTGTGGAGCTGTGGGCAATTTCTAGCAATCGT 432
508 GCACGTAGAGGCTGCATATACACAGCTAACTGAAAGAGACATCATATTCTGAACAGCAA 567
433 GCAGTAGAGGATTAATAAGATAGTAAACGAGGAGCTGATTTCTGTTATCAGAACAGAA 492
568 CTTGTGAGCTGTGCTCAGCATTCAATTAATCTTGGATGCAATGAGGTTTSCCTTCCCAA 627
493 TTGGTGGATTGTGACTGGAT---ATAACCAAGGATGCAATGGAGGACTTATGGACTAT 549
628 GCCTTTGAATACGTTAAGTACAAATGAGGATGAGTTGAAGCATGCACTGGGCTTGGTGCTCA 807
670 TAGGAAGATGTTCTGCTGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 729
808 GTTAGCGTTGCAATTTGAGGTTGTGAAGGTTTCAATCTGTACAAAGAGGTTGTATACAGC 867
730 GTGAGCGTTGCTATTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 789
868 AGTGACACTGTGGAAGATCAATGATGATGATGATGATGATGATGATGATGATGATGAT 927
790 ACGGGAGATGTGGCAGACAC-----TAGATCATGTTGTGTTGCTGTTGGGTAT 840
928 GGAGTCGAGAGCGGAGTTCTTATTGGCTCATCAAGAACTATGGGTTACAAATTTGGGT 987
841 GGAACAGCTAAAGTTCATGATTTGATGATGATGATGATGATGATGATGATGATGATGAT 900
988 GACAATGGCTACTTTAAGATGGA 1011
901 GAGGATGGCTACATCAGATTAGAA 924

RESULT 9

US-08-813-591-1
; Sequence 1, Application US/08813591
; Patent No. 5824534
; GENERAL INFORMATION:
; APPLICANT: ASANO, MINAO
; APPLICANT: KAWAI, MISAKO
; APPLICANT: MIWA, TETSUYA
; APPLICANT: NIO, NORIKI
; TITLE OF INVENTION: AMINOPEPTIDASE GX, AND A METHOD OF
; TITLE OF INVENTION: HYDROLYSING A PROTEIN WITH THE SAME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: P. C.
; ADDRESS: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/813,591
FILING DATE: 07-MAR-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 051848/1996
FILING DATE: 08-MAR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 030458/1997
FILING DATE: 14-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-845-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1056 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Glycine max
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1056
OTHER INFORMATION: /note= "METHOD OF DETERMINING THE
OTHER INFORMATION: CHARACTERISTICS: P"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..9
OTHER INFORMATION: /note= "INSERTION SEQUENCE, METHOD
OTHER INFORMATION: OF DETERMINING THE CHARACTERISTICS: E"
US-08-813-591-1

Query Match 14.7%; Score 157.6; DB 1; Length 1056;
Best Local Similarity 52.7%; Pred. No. 1.2e-38;
Matches 455; Conservative 0; Mismatches 379; Indels 30; Gaps 4;

QY 166 TTCACAGCTTTCGCGCCGAGGTACGGGAGAGCTACGGATCGGAGGAGGAGATCAAGAG 225
Db 73 TACGAGCAGTGGCTCGTGAAGCAGCGGAGGTTGTACAAACGCTCGCGGAGAGAGAG 132
QY 226 AGTTTCGGGATCTTCGTGGAGATCTAGCGTTTATCCGTTCCACTAA---TCGGAAGAT 282
Db 133 CGCTTCCAGATCTTCAAGGACCACTCGGATTCATCGACCACTCCGCGGAGGAC 192
QY 283 CTGTCTATACCTTAGGAATCAACCAATTCGCGACCTGACCTGGGAGGAAATTCGGACC 342
Db 193 CGAACCTACAGCTCGGAGTGAACCGGTTCTGCTGATCTCACCACGAGGATACAGGGCC 252
QY 343 AATCGCTTGTGGCGGCGAGAA-----CTGCTCGGAGCTGCGCATGGA 387
Db 253 AAGTACTTGGGAACCAAGATCGATCCCAACCGGAGGCTCGGAAGAGACCCCGAGCAACCGC 312
QY 388 AACCAACCGGTTTGTGATGGCTCTTCTGTAAACGAGGATTTGGAGGAGCAAGGATA 447
Db 313 TAGCGCCACGCTGTGGGAGCAATTTGCTGATTCGGTTGATTTGGAGGAAGGAGTGCT 372
QY 448 GTGAGCCCTGTAAAGGACCAAGGAGCTGTGGATCTTCTGCTGAGCTTTCAGTACTACTGGA 507
Db 373 GTTCTCTCTGTCAAAAGACCAAGGAGGCTGTGGGAGCTGTGGGCAATTTCTAGCAATCGT 432
QY 508 GCACGTAGAGGCTGCATATACACAGCTAACTGAAAGAGCAGCATCATATTCTGAACAGCAA 567
Db 433 GCAGTAGAGGATTAATAAGATAGTAAACGAGGAGCTGATTTCTGTTATCAGAACAGAA 492
QY 568 CTTGTGAGCTGTGCTCAGCATTCATTAATCTTGGATGCAATGAGGTTTSCCTTCCCAA 627
Db 493 TTGGTGGATTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 549

628	GCCTTTGAATACGTTTAAGTACAATAGGAGGCATGACACAGAACAGACTTATTCATACCTT	687
550	GCAATTTGAGTTTCATAATCAACAATGGCGCATTCGATTTCTGATGAGGATTAACCATACCGT	609
688	GGTGTCAATGGTATCTGCAACTTCAACGACGAGAGAACTGGTGTCAAGGTCAATTTGATTCG	747
610	GGTGTGATGGTAGATGGCACATATAGGAAAAATGCTAAAGTCGTTCTTATTTGATGAC	669
748	ATAAATCATCACCCCTGGGTCTGAGGATGAGTTGAAGCATGCAGTGGGCTTTGGTGCCTCCA	807
670	TACGAAGATGTTCTGCGCTATGATGAGTTAGCCCTTGAAAAGGCCGTTGCAAAATCAGCCC	729
808	GTTAGCGTTGCATTTGAGGTTGTGAAGTTTCAATCTGTACAAGAAAGGTGTAATACAGC	867
730	GTGAGCGTTTGCTATTGAAGAGGGGGCAGGGAAATTTCAATTTATGTATCTGGTGTATTTC	789
868	ACTGACACCTGTGGAAGAGATCCCAATGGATGTGAACCAACGACAGTCTTTGCCGTCGGTTAT	927
790	ACGGGAGATGTGGCACAGCAC-----TAGATCATGGTGTGCGGCTGTGGGTAT	840
928	GGAGTCGAGACGGGATTCCTTATTGGCTCATCAAGAACTCATGGGGTACAATAATGGGGT	987
841	GGAAAGACTTAAGGTCATGATTTATGGATCGTAGGAAATCATGGGGTTCTAGCTCGGGA	900
988	GACAAATGGCTACTTTAAGATGGAA	1011
901	GAGGATGGCTACATCAGATTAGAA	924

```

RESULT 10
US-09-325-932A-112
; Sequence 112, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: death and their use in the modification of forestry plant develop
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 112
; LENGTH: 522
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-09-325-932A-112

```

Query Match	14.5%;	Score 155;	DB 4;	Length 522;
Best Local Similarity	63.8%;	Pred. No. 5.1e-38;		
Matches 252;	Conservative	0;	Mismatches 140;	Indels 3; Gaps 1;
Qy	116	TTGAGCTCGCATCTCTCGGCGTCTTCGGCAGTTTCGCCACAGCCTTCACACTTCGCACACGGT	175	
Db	128	TTGAGTCTCTCATCTCTCAACACCGTCGGCCACACCCGTCGCGCCCTCTCTCTTCGTAGACT	187	
Qy	176	TGCCCCGCMGGTACGGGAAGACTACGGATCGGAGGAGGAGATCAAGAAGAGGTTTCGGGA	235	
Db	188	TGCCCCGGGGCAGCGGAAGACTTACAAGACAGCGGAGGAGATCAAGTTGAGGTTTCGATA	247	
Qy	236	TCTTCGTGGAGAATCTAGCGTTTATCCGGTCCACTAATCGAAGGATCTGTGATATACC	295	
Db	248	ATTACAGGAGAACTCTCAAGTCAATTCGATCCACCAAGGAGGCTTGGCTTTACATCTC	307	
Qy	296	TAGGAATCAACCAATTCCGCGACCTGCCTGGGAGNAATTCGGAGCAATCCGCTTGGTG	355	
Db	308	TCGCTGTTAATCAGTATCTCTACCTGAGCTGGGAGGAGTTCAAGACGACAGCTGGGAG	367	
Qy	356	CGCGCGAGAATCTGCTCGCGGACTCGCATCGTAAACCAACCGGTTTGTGATCGCGTGCTTC	415	
Db	368	CTTCTCAAGACTGTCTCTCCACCAACCAAGGCGAGCACAAAGCTCAACGACGATGTTCTTC	427	

416	Qy	CTGTAAACAGAGGATTTGGGGAGCAACGGATATGAG---CCCTGTAAAGGACCCAGGAA	472
428	Db	CCGAACCAAGAACCTGGGAGAGAAAGGCGATTTGATGCCCCAGTTAAAGATCAAGCGG	487
473	Qy	GCTGTGGATCTTGTGGACTTTCAGTACTACTGGA	507
488	Db	CTTGTGGATCTTGTGGAGTTTCAGCGCAACTGGA	522

RESULT 11
US-08-883-526-2
Sequence 2, Application US/08883526
Patent No. 6033893
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Purvi
TITLE OF INVENTION: NEW HUMAN CATHESPIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/883,526
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0331 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-845-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1366 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: THYMNOT02
CLONE: 347021
US-08-883-526-2

	Query Match	14.1%;	Score 150.6;	DB 3;	Length 1366;
	Best Local Similarity	51.9%;	Pred. No. 2e-36;		
	Matches 392;	Conservative	0;	Mismatches 354;	Indels 9; Gaps 2;
Qy	279	GGATCTGTCGTATACCCCTAGAAATCAACCAATTCGCCGACCTGACCTGGGAGGAATTCGG	338		
Db	266	GAACCTGGCTTCACAAATGGCCATGAATGCTTTGGTGACATGACCAATGAAGAAATTCAG	325		
Qy	339	GACCAATCGCCTTTGGTGGCGGCGAGAACTGCTCGGGGACTGCCGCTGGAACCAACCGGTT	398		
Db	326	GCAGATCATGGGTTGCTTTCCGAAACCGAAATTCAGGAGGGGAAGTGTTCGTGAGCC	385		
Qy	399	TGTCGATGGCGTCTTCTCTTAACGAGGGGATTCGAGGGAGCAAGGGGATATGTGAGCCCTGT	458		

386 TCTGTTTCTTGTATCTCCCAATCTGTGGATTGGAGAAAGAGGCTACGTGACGCCAGT 445
459 AAGAGACCAAGAGCTGTGGATCTTGTGACTTTCAGTACTACTAGCACTAGAGGC 518
446 GAAGAATCAAGAACAGTGTGGTCTTGTGGCTTTTGTGGCTTTTGTGGCTTTTGTGG 505
519 TGCATATACACAGCTAACTGGAAGAGACACATCTATCTGAACAGCACTTTGGACTG 578
506 ACAGATGTTCCGGAACCTGGGAACCTTGTCTCACTGAGCGAGCAGATCTGGTGACTG 565
579 TGCCTCAGCATCAATACTTTGGATGCAATGGAGGTTTGCCTCCCAAGCCTTTGAATA 638
566 TTGCGCTCTCAAGCAATCAGGCTGCAATGGTCTTATGGCTTAGGCTTCCAGTA 625
639 CGTTAAGTCAATGAGGATCGACAGAACAGACTTATCATACCTTGTGTCAATGG 698
626 TGTCAAGGAGACGAGGCTGCACTCTGAGGAATCTTATCATATGATGAGTGA 685
699 TATCTCAACTTCAAGCAGAGAGAAATGTTGGTCAAGTCAATGATTCGATTAACATCAC 758
686 AATCTGTAAGTACAGACCTGAGAAATCTTGTCTAATGACACTGGCTTCACAATGGTCG 745
759 CTTGGTGTCTGAGTGAATGAGCATGAGCTGCTGGCTTGGTGGCTGCTGAGTGGTGC 818
746 ACTGGAAGAGAGAGGCTGATGAAGCAGTTCGCACTGTGGGCTCCATCTCGTTCG 805
819 ATTGAGGTGTGAA---AGGTTTCAATCTGTACAGAAAGGTGTATACAGCAGTGACAC 875
806 TATGATGAGGCAATTCCTGCTTCCAGTCTTACAAATCAGGCAATTTATTTGAACCCAGA 865
876 CTGTGAAGAGATCCAAATGATGTGAACCAAGCTTCTTGGCTGGTGGTATGAGTGA 935
866 CTGACAGCAAAACCTGATCATGATGATGATGATGATGATGATGATGATGATGATG 925
936 GAGCGGATTCCT-----TATTGGCTCATCAAGAACTCATGGGTCACAAATGGGGTGA 989
926 AATTCGAATACAGCAAGTATTGGCTCTGCAAAACAGCTGGGTCAGATGGGCTC 985
990 CAATGGCTATTGAATGAACTCGGCAAGACA 1024
986 GAATGGCTATGTAATAATAGCCAAAGACAAAGAACA 1020

RESULT 12

US-09-519-283-2

Sequence 2, Application US/09519283

Patent No. 6620606

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

Guegler, Karl J.

Corley, Neil C.

Shah, Purvi

TITLE OF INVENTION: NEW HUMAN CATHEPSIN

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/519,283

FILING DATE: 07-Mar-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/883,526

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PP-0331 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1366 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: THYMOT02

CLONE: 347021

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-519-283-2

Query Match

Best Local Similarity 14.1%; Score 150.6; DB 4; Length 1366;

Matches 392; Conservativity 0; Mismatches 354; Indels 9; Gaps 2;

Qy 279 GGATCTCTCTATACCTAGGAATCAACCAATTCGCGACCTGACCTGGGAGGAATTCG 338
Db 266 GAAACTTGGCTTCAAAATGCCATGATGCTTTTGTGACATGACCAATGAAGATTCAG 325
Qy 339 GACCAATCGCTTGTGCGCGCAGAACTCTCTCGGCGACTGCGCATGGAACACACCGTT 398
Db 326 GCAGATGATGGTGTCTTCGAAACCAAGAAATTCAGAAAGGAAAGTGTTCGCTGAGCC 385
Qy 399 TGTGATGGGTGCTTCTCTGACGAGGATTCGAGGAGGAGGAGGAGGAGGAGGAGG 458
Db 386 TCTGTTTCTTGTATCTTCCAAATCTGTGAGATGGAGAAAGAAAGGCTACGTGAGCCAGT 445
Qy 459 AAAGGACCAAGGAGCTGTGATCTTCTGCGACTTTTCAGTACTACTGAGCAGCTAGAGGC 518
Db 446 GAGAAATCAGAAACAGTGTGTTCTTCTTGGGCTTTTGTGCGACTGCTGTCTTGAAG 505
Qy 519 TGCATATACAGCTAATCGAAGAGCAGACATCATCTGAAACAGCAACTTGTGGACTG 578
Db 506 ACAGATGTTCGGAAGAACTGGGAAACTTGTCTCACTGAGGAGCAGAAATCTGTGGACTG 565
Qy 579 TGCCTCAGCATTCATAAATTTGGATCAATGGAGGTTTGCCTTTCCCAAGCCTTTGAATA 638
Db 566 TTGCGCTCTCAAGCAATCAGGCTGCAATGGTGGCTTCATGCTAGGSCCTTCCAGTA 625
Qy 639 CGTTAAGTACAAATGAGGAGATCGACACAGAACAGACTTTATCCATACCTTGTGTGATGG 698
Db 626 TGTCAAGGAGAAACGAGGCTGGAATCTGAGGAATCTCTATCCATATGATGAGTGAATGA 685
Qy 699 TATCTGCAACTTCAAGCAGGAGATGTTGGTGTCAAGGTCAATGATTGATTAACATCAC 758
Db 686 AATCTGTAAGTACAGACTGAGAAATCTGTGTAATGACACTGGCTTCAATGAGTGG 745
Qy 759 CTTGGGTCTCAGGATGAGTTGAAGCAGTGGGCTTGGTGGCTTGGTGGCTTGGTGGCT 818
Db 746 ACCTGGAAGAGAGAGGCTGATGAAAGCAGTCCCAACTGTGGGGGCCATCTCCGTTGC 805
Qy 819 ATTTGAGTTGTGAA---AGGTTTCAATCTGTCAAGAAAGGTGTATACAGCAGTGACAC 875
Db 806 TATGATGAGGCAATTCGCTCTTCCAGTCTTCAAAATCAGGCAATTTATTTGAACCCAGA 865
Qy 876 CTGTGGAAGAGATCCAAATGATGTGAACCAAGCAGTTCCTTCCCGCTGGTGTGAGTGA 935
Db 866 CTGACAGCAGCAAAACCTGAGTATGTTGTTGCTGTTGGCTTGGCTTGGCTTGGAGGAGC 925
Qy 936 GAGCGGATTCCT-----TATTGGCTCATCAAGAACTCATGGGTCACAAATGGGGTGA 989
Db 926 AATTCGAATACAGCAAGTATTGGCTCTGCAAAACAGCTGGGTCAGATGGGCTC 985
Qy 990 CAATGGCTATTGAATGAACTCGGCAAGACA 1024

Db 986 GAATGGCTATGTAAATAATAGCCAAAGACAAAGAACA 1020
|||||

RESULT 13

US-08-821-994-61
; Sequence 61, Application US/08821994A
; Patent No. 6228643

GENERAL INFORMATION:

; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jepson, Ian

; TITLE OF INVENTION: Promoters

; FILE REFERENCE: PPD 50108

; CURRENT APPLICATION NUMBER: US/08/821,994A

; CURRENT FILING DATE: 1997-03-22

; EARLIER APPLICATION NUMBER: PCT/GB97/00729

; EARLIER FILING DATE: 1997-03-18

; EARLIER APPLICATION NUMBER: GB 9606062.9

; EARLIER FILING DATE: 1996-03-22

; NUMBER OF SEQ ID NOS: 89

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 61

; LENGTH: 1390

; TYPE: DNA

; ORGANISM: Brassica napus

US-08-821-994-61

Query Match 13.8%; Score 147.4; DB 3; Length 1390;

Best Local Similarity 55.2%; Pred. No. 2e-35;

Matches 336; Conservative 0; Mismatches 261; Indels 12; Gaps 2;

QY 402 CGATGGCGTGCTTCCTGTAAACGAGGATTTGAGGGAGCAAGGGATAGTGAGCCCTGTAA 461

Db 467 CGACGTGGAGTTCCGGAGACGGTTGACTGAGACAGAGAGGACCGTTAATGCCATTAA 526

QY 462 GGACCAAGAGCTGTGGATCTTCTGTGACCTTTTCAGTACTACTGAGGACCTAGAGGCTGC 521

Db 527 AAACCAAGACTTTCGCGAAGTTTGTGGCGTTTTCACACAGCTGCAGCAGTAGAAGGCAT 586

QY 522 ATATACACAGCTAACTGGAAGAGACACATATTCTGAACAGCAACTTGTGGACTGTGC 581

Db 587 AAACAGATCGTAACAGGAGAGCTCATATCTCTGTCGACACAGAACTTGTGACTGCGA 646

QY 582 CTCAGACTTCAATTTTGGATGCAATGGAGGTTTGCCTTCCCAAGCCTTTGAATAGT 641

Db 647 CAATCAT---ACAACCAAGGCTGTAAACGGCGTCTAATGGATTATGCTTTCAATTAT 703

QY 642 TAAGTACAAATGGAGCATCGACACAGACAGACTTATCCATACCTTGGTGTCAATGGTAT 701

Db 704 CATGAAAACCGCGGATTAACACCGAGCAAGACTATCCCTTACCACGGAACCAATGGCAA 763

QY 702 CTGCAACTTCAAGCAGGAGAAATTTGGTGTCAAGGTCATTGATTGATTAACATCACCT 761

Db 764 ATGCAACTCTTTACTTAAATTCGAGAGTTGTGACTATCGATGATACGAGATGTTCC 823

QY 762 GGGTGTGAGGATGAGTTGAAGCATGACGTGGGCTTGGTGGTCCAGTTAGCGTTGCATT 821

Db 824 TAGTAAGATGAAACCGCGTTTGAAGAGAGCAGTTTCGTACCGAGCTGTGAGTTGCTAT 883

QY 822 TGAGTTGTGAAAGTTTCAATCTGTACAGAAAGGTTATACAGCAGTGCACACCTGTGG 881

Db 884 TGATGCTGTGGAAGAGCTTCCAACTTACCATCTGGATCTTCTACTGGAAGTGTGG 943

QY 882 AAGAGATCCAATGGATGTAACCAACGAGTTCTTGGCGTGGTTATGGAATCGAGGACGG 941

Db 944 -----TAGCACTATGGATCAAGCTGTTGTGGCGGTTGGTTATGGATCAGAGAACGG 994

QY 942 GATTCCTTATGGCTCATCAAGAACTCATGGGTGACAAATTTGGGTGACAAATGGCTACTT 1001

Db 995 TGTGACTATTGGATTGTACGTAACCTTTGGGTGACAAAGCTGGGAGAGGATGTTACAT 1054

QY 1002 TAAGATGGA 1010

Db 1055 TAGGATGGA 1063
|||||

RESULT 14

US-08-821-994-63

; Sequence 63, Application US/08821994A

; Patent No. 6228643

GENERAL INFORMATION:

; APPLICANT: Greenland, Andrew J

; APPLICANT: Thomas, Didier RP

; APPLICANT: Jepson, Ian

; TITLE OF INVENTION: Promoters

; FILE REFERENCE: PPD 50108

; CURRENT APPLICATION NUMBER: US/08/821,994A

; CURRENT FILING DATE: 1997-03-22

; EARLIER APPLICATION NUMBER: PCT/GB97/00729

; EARLIER FILING DATE: 1997-03-18

; EARLIER APPLICATION NUMBER: GB 9606062.9

; EARLIER FILING DATE: 1996-03-22

; NUMBER OF SEQ ID NOS: 89

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 63

; LENGTH: 1441

; TYPE: DNA

; ORGANISM: Brassica napus

US-08-821-994-63

Query Match 13.8%; Score 147.4; DB 3; Length 1441;

Best Local Similarity 55.2%; Pred. No. 2.1e-35;

Matches 336; Conservative 0; Mismatches 261; Indels 12; Gaps 2;

QY 402 CGATGGCGTGCTTCCTGTAAACGAGGATTTGAGGGAGCAAGGGATAGTGAGCCCTGTAA 461

Db 453 CGACGTGGAGTTCCGGAGACGGTTGACTGAGACAGAAAGGAGCGGTTAATGCCATTAA 512

QY 462 GGACCAAGAGAGCTGTGGATCTTCTGTGACTTTCAGTACTACTGAGGACCTAGAGGCTGC 521

Db 513 AAACCAAGAGATCTTTCGGAAGTTTGTGGCGTTTTCACACAGCTGCAGCAGTAGAAGGCAT 572

QY 522 ATATACACAGCTAACTTGGAAAGAGACATATTCTGAACAGCAACTTGTGGACTGTGC 581

Db 573 AAACAGATCGTAAACAGGAGAGCTCATATCTCTGTCGACACAGAACTTGTGACTGCGA 632

QY 582 CTCAGACTTCAATAACTTTTGGATGCAATGGAGGTTTGCCTTCCCAAGCCTTTGAATAGT 641

Db 633 CAATCAT---ACAACCAAGGCTGTAAACGGCGTCTAATGGATTATGCTTTTCAATTAT 689

QY 642 TAAGTACAAATGGAGCATCGACACAGACAGACTTATCCATACCTTGGTGTCAATGGTAT 701

Db 690 CATGAAAACCGCGGATTAACACCGAGCAAGACTATCCCTTACCACGGAACCAATGGCAA 749

QY 702 CTGCAACTTCAAGCAGGAGAAATTTGGTGTCAAGGTCATTGATTTCGATAAACATCACCT 761

Db 750 ATGCAACTCTTTACTTAAATTCGAGAGTTGTGACTATCGATGATACGAGATGTTCC 809

QY 762 GGGTGTGAGGATGAGTTGAAGCATGACGTGGGCTTGGTGGTCCAGTTAGCGTTGCATT 821

Db 810 TAGTAAGATGAAACCGCGTTTGAAGAGAGCAGTTTCGTACCGAGCTGTGAGTTGCTAT 869

QY 822 TGAGGTTGTGAAAGTTTCAATCTGTACAGAAAGGTTATACAGCAGTGCACACCTGTGG 881

Db 870 TGATGCTGTGGAAGAGCTTCCAACTTACCAATCTGGAATCTTCACTGGAAGTGTGG 929

QY 882 AAGAGATCCAATGGATGTAACCAACGAGTTCTTGGCGTGGTTATGGAATCGAGGACGG 941

Db 930 -----TAGCACTATGGATCAAGCTGTTGTGGCGGTTGGTTATGGATCAGAGAACGG 980

QY 942 GATTCCTTATGGCTCATCAAGAACTCATGGGTGACAAATTTGGGTGACAAATGGCTACTT 1001

Db 981 TGTGACTATTGGATTGTACGTAACCTTTGGGTGACAAAGCTGGGAGAGGATGTTACAT 1040

QY 1002 TAAGATGGA 1010

Db 1041 TAGGATGGA 1049

Db 878 TAGGATGGA 886

Search completed: March 29, 2004, 21:10:30
Job time : 83 secs

RESULT 15
US-08-821-994-86
; Sequence 86, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jenson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; EARLIER FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 86
; LENGTH: 1102
; TYPE: DNA
; ORGANISM: Brassica napus
US-08-821-994-86

Query Match 13.6%; Score 145.8; DB 3; Length 1102;
Best Local Similarity 55.0%; Pred. No. 5.5e-35;
Matches 335; Conservative 0; Mismatches 262; Indels 12; Gaps 2;
QY 402 CGATGCGCTGCTTCCTGTAACGAGGATGGAGGGAGCAAGGGATAGTGAGCCCTGTAAA 461
Db 290 CGACGTGGAGGTTCCGGTACCGTTGACTGGAGACAGAAAGGAGCGCTTAATGCCATTAA 349
QY 462 GGACCAAGGAAGCTGTGGATCTTGTGGACTTTTCACTACTGAGGACACTAGAGGCTGC 521
Db 350 AGACCAAGGAACCTTGGGGAAGTTTGGCGGTTTTCAACAGCTGCAGCAGTAGAAGGTAT 409
QY 522 ATATACACAGCTAATGGAAGAGACATCATATCTGACAGCAACTTGTGACTGTGC 581
Db 410 AAACAAGATCGTAACAGGAGAACTCGTATCTTTGTCGGAACAAGAACTTGTGACTGGCA 469
QY 582 CTCAGCATTCATAATCTTTGGATGCAATGGAGGTTTGGCCTTCCCAAGCCTTTGAATAGCT 641
Db 470 CAATCGT---ACAACCAAGGCTGTAAACGGCGTCTAATGGATTATGCTTTTCAATTAT 526
QY 642 TAAGTCAATGGAGGCAATCGACACAGAACTTATCCATACCTTGGTGTCAATGGTAT 701
Db 527 AATGAAAAACGGCGGATTAAACACCGAGAAAGACTATCCTTACCACGGAACCAATGGCAA 586
QY 702 CTGCAACTTCAAGCAGGAGAACTGTTGGTGTCAAGGTCATTGATTCGATAAACATCACCT 761
Db 587 ATGCAACTCTTACTTAAGATTCAGAGTTGTAACATTCGATGGATACGAGATGTTCC 646
QY 762 GGGTGTGAGGATGAGTGAAGCATGCACTGGGCTTGGTGGTCCAGTTAGCGTTGCATT 821
Db 647 TAGTAAAGATGAACCGCGTTGAAGAGAGCAGTTTCATACAGCCTGTGAGTGTGCTAT 706
QY 822 TGAAGTTGTGAAGTTTCAATCTGTACAGAAAGGTGTATACAGAGTGACACCTGTGG 881
Db 707 TGATGCTGTGAAGAGCTTTCCAACTTACCAATCTGGAATCTTCACTGGAAAGTGTGG 766
QY 882 AAGAGATCCAATGGATGTGAACCAACGAGTTCTTTCGCGTCCGTTATGGAGTCGAGGCGG 941
Db 767 -----TACGAATATGATCACGCTGTGGTGGCGGTTGGTTATGGGTACAGAGACGG 817
QY 942 GATTCCTTATGGCTCATCAAGAACTCATGGGTAACAATTTGGGTGACAAATGGCTACTT 1001
Db 818 CGTTGACTATGGAATTGTAAGTAACTTTGGGGTACACGTTGGGGAGAGATGGTTACAT 877
QY 1002 TAAGATGGA 1010

Db	61	GC	CGGTTTCGAAGAAAGCAATCCAAATCCGGTCCGTTTACAAAGAGCCTGACTCGATTGAG	120
Qy	121	CT	GCACATCCCTCGCGCTCCTTTGGCAGTTGGCCACGCTTCCACTTCGCA	180
Db	121	CT	GCACATCCCTCGCGCTCCTTTGGCAGTTGGCCACGCTTCCACTTCGCA	180
Qy	181	CG	CAGGTACCGGAGAGACTACGGATCGAGGAGGAGATCAAGAGAGGTTCCGGATCTTC	240
Db	181	CG	CAGGTACCGGAGAGACTACGGATCGAGGAGGAGATCAAGAGAGGTTCCGGATCTTC	240
Qy	241	GT	GAGAACTTAGCGTTTATCCGCTCCACTAATCGGAAGGATCTGTCGTATACCC	300
Db	241	GT	GAGAACTTAGCGTTTATCCGCTCCACTAATCGGAAGGATCTGTCGTATACCC	300
Qy	301	AT	CAACCAATTCGCGACCTGACCTGGAGGAATTCGGACCAATCCCTTGGTGGCGG	360
Db	301	AT	CAACCAATTCGCGACCTGACCTGGAGGAATTCGGACCAATCCCTTGGTGGCGG	360
Qy	361	CAG	AATCTGCTCGGCACTGGCAGTGGAAACACCGGTTTTCGATGCGCTTCCTGT	420
Db	361	CAG	AATCTGCTCGGCACTGGCAGTGGAAACACCGGTTTTCGATGCGCTTCCTGT	420
Qy	421	AC	GAGGATTTGGAGGAGCAAGGATAGTGAGCCCTGTAAGGACCAAGGACGTGGGA	480
Db	421	AC	GAGGATTTGGAGGAGCAAGGATAGTGAGCCCTGTAAGGACCAAGGACGTGGGA	480
Qy	481	TC	TGCTGCTGGACTTTTCAGTACTACTCTGAGACACTAGAGGCTGCATATACAC	540
Db	481	TC	TGCTGCTGGACTTTTCAGTACTACTCTGAGACACTAGAGGCTGCATATACAC	540
Qy	541	AAG	AGCACATCATTTATCTGAACAGCAACTTGTGCACTGCGCTCAGCAATTC	600
Db	541	AAG	AGCACATCATTTATCTGAACAGCAACTTGTGCACTGCGCTCAGCAATTC	600
Qy	601	GG	ATGCAATGGAGGTTTGGCTTCCCAAGCCCTTGAATACGTTAAGTACAATG	660
Db	601	GG	ATGCAATGGAGGTTTGGCTTCCCAAGCCCTTGAATACGTTAAGTACAATG	660
Qy	661	GAC	CAGAAACGACTTATCCATCTTGGTGTCAATGGTATCTGCAACTTCAAG	720
Db	661	GAC	CAGAAACGACTTATCCATCTTGGTGTCAATGGTATCTGCAACTTCAAG	720
Qy	721	AAT	GTCGTGTCAAGGTCATTGATTCGATAAACATCACCCCTGGGTGTGAGGAT	780
Db	721	AAT	GTCGTGTCAAGGTCATTGATTCGATAAACATCACCCCTGGGTGTGAGGAT	780
Qy	781	AAG	CAATCGAGTGGGCTTGGTGGTCCGATAGCGTTGCATTTGAGGTTTGAA	840
Db	781	AAG	CAATCGAGTGGGCTTGGTGGTCCGATAGCGTTGCATTTGAGGTTTGAA	840
Qy	841	AAT	CTGTACAAGAAAGGTGTATACAGAGTGCACCTGTGGAAGAGATCCAA	900
Db	841	AAT	CTGTACAAGAAAGGTGTATACAGAGTGCACCTGTGGAAGAGATCCAA	900
Qy	901	AAC	CACGACTTTCGCGCTTATGAGTTCGAGACGGGATTCCTTATTTGGCT	960
Db	901	AAC	CACGACTTTCGCGCTTATGAGTTCGAGACGGGATTCCTTATTTGGCT	960
Qy	961	AAG	AACATCATGGGGTACAAATTGGGGTGCAATGGCTACTTTAGATGGAACT	1020
Db	961	ARG	AACTCATGGGGTACAAATTGGGGTGCAATGGCTACTTTAGATGGAACT	1020
Qy	1021	AAC	ATGTTGGTGTGCAACTTCGCACTTTATCCCAATGTCGCTGTGTAG	1071
Db	1021	AAC	ATGTTGGTGTGCAACTTCGCACTTTATCCCAATGTCGCTGTGTGTAG	1071

RESULT 2

RESULT 2
US-10-425-114-156

US-10-423-114-136
; Sequence 156, Application US/10425114

; Publication No. US20040034888A1

; GENERAL INFORMATION:

APPLICANT: Liu, Jingdong

[illegible]

QY 543 GAGCATCATTTATCTGAACAGCAACTTGTGGACTGTGGCTCAGCATTTCAATTAATTTGG 602
DB 612 GCGCATCTCTCTCTGTGAGCAACAGCTTATTGACTGTGGTTTGATTTCAACAATTTGG 671
QY 603 ATGCAATGAGGTTTGGCTTCCCAAGCCCTTGAATACGTTAAGTACAAATGAGGATCGA 662
DB 672 ATGCAACGAGGAGGCTTCCATCCAGCCCTTGAATACATCAATACATGATGAGGCTTGA 731
QY 663 CACAGAACAGATTTATCATCTTGGTGTGATGATGATGATGATGATGATGATGATGATG 722
DB 732 CACTGAGGAATCTTACCCCTTACCAAGGTGTCAATGGAATCTGCAAGTTTAAAGAAATG 791
QY 723 TGTGGTGTCAAGGTCAATTTGATTCATATAAATCAATCAACCTGGGTGCTGAGGATGATGAA 782
DB 792 TGTGGAGTCAAGTTTGGACTCGTTTAAATCAATCAACCTGGGTGCTGAGGATGATGAA 851
QY 783 GATGCAAGTGGCTTGGTGGCTTGGTGGCTTGGTGGCTTGGTGGCTTGGTGGCTTGGTGG 842
DB 852 GATGCTGT 911
QY 843 TCTGTACAGAAAGGTGTATACAGCAGTGCACCTGTGGAAGATCCAAATGGATGAA 902
DB 912 GCTGTACAGAGCGAGTTTACATAGCGCAATTTGTGAACTACACCGATGGATGAA 971
QY 903 CCACGAGTCTTGTGGCTTGGTGGCTTGGTGGCTTGGTGGCTTGGTGGCTTGGTGGCT 962
DB 972 CCACGCTGT 1031
QY 963 GAACATCATGGGTACAAATTTGGGTGACATGGCTTAAATGATGGAATCGGCAAGAA 1022
DB 1032 GAACATCATGGGTGCTGACTGGGTGATGAGGGTTACTTCAAGATGGAATGGGCAAGAA 1091
QY 1023 CATGT 1064
DB 1092 CATGT 1133

RESULT 6
US-10-425-114-5145
; Sequence 5145, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-2153313B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 5145
; LENGTH: 1367
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700444833_FLI
US-10-425-114-5145

Query Match 51.7%; Score 553.8; DB 12; Length 1367;
Best Local Similarity 72.9%; Pred. No. 1.6e-175;
Matches 743; Conservative 0; Mismatches 267; Indels 9; Gaps 2;

QY 55 GCGCTCGCGGTTTCGAAGAAGACAATCCAAATCCGGTCCGTTACACAAAGCCCTGACT-- 112
DB 1 GCCAACTCGGCTTTCGGGACTCCAAACCGATCCGCTGTACCGACCGCGGCGCTCC 60
QY 113 -CGATTAGCTGCGCATCTCGGCTGCTTGGCAGTTGCGCCACGCTTCCATTCGCA 171
DB 61 GCGCTGAGTCCACGGTCTTCGCGGCTCGCGCGCACCGCGAGCGGCTCGCTTCGCA 120

RESULT 7
US-10-219-220-107
; Sequence 107, Application US/10219220
; Publication No. US2003008274A1
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: Death and their use in the modification of plant development
; FILE REFERENCE: 11000.1022c1
; CURRENT APPLICATION NUMBER: US/10/219,220

QY 172 CGGTTCCGCGCAGGTACCGGAAGAGCTACGATCGAGGAGGAGATCAAGAAAGAGTTTC 231
DB 121 CGCTTCCGCGTCAAGTACGGAAGAGCTACGAGAGCGCGGAGGTCCATAAGCGGTTTC 180
QY 232 GGGATCTTGTGAGAACTTAGAGTTTATCCGGTCCACTAATCGGAAGGATCTGTGAT 291
DB 181 AGGATCTTCTCCGAGAGCTTCCAGCTGGTCCGCTCCCAACCGCAAGAGGCTTCTCTAC 240
QY 292 ACCTTAGGAATCAACCAATTCGCGGACCTGACCTGGGAGGAATTCGGGACCAATCGCTT 351
DB 241 CGCTCGGATCAACCGCTTCCGCGACATGAGCTGGAGGAGTTCCGTGCGACCGGCTC 300
QY 352 GGTGCGCGCAGAACTGCTCGGCGAGCTGGCGATGGAACCCACCGTT-----TGTGAT 405
DB 301 GCGCGACCCAGAACTGCTCGGCGACGCTTACCGGCAACCCACCGATGCGCGCGCGCC 360
QY 406 GGGTGTCTTCTCTAAGAGGAGATTCGAGGAGCAAGGATAGTAGCCCTGTAAAGGAC 465
DB 361 GTTGGCTGCGGAGAGCAAGACTGGAGGAGGATGGGATTTGAGCCAGTGAANAAC 420
QY 466 CAAGGAAGCTGTGGATCTTGTGGACTTTCAGTACTACTGGAGCACTAGAGGCTGCATAT 525
DB 421 CAGGCGCACTGTGGATCATGCTGGACCTTTCAGCACTACTGGTGCATTTGAGGCTGCATAT 480
QY 526 ACACAGCTAATCGGAAGAGCACTCATTTATCTGAACAGCAACTTGTGAGCTGTGCTCA 585
DB 481 ACCAGGCACTGTGCAAGCCATCTCTCTCTGAGCAACAGCTTGTGACTGTGGTTTT 540
QY 586 GCATTAATAACTTTGATGCAATGAGGTTTGGCTTCCAAAGCCCTTGAATAAGTTAAG 645
DB 541 GCATTAACAATTTTCGGATGCAAGGAGGCTTCCATCCAGGCTTTCATATACATCAAA 600
QY 646 TACAATGGAGGATCGACACAGCAAGACTTATCCATACCTTGGTGTCAATGATGATCTGC 705
DB 601 TACAATGGTGGCTTGCACACTGAGGAATCTTACCTTACCAAGGCTGCAATGGAATCTGC 660
QY 706 AACTTCAAGCAGGAGAAATTTGGTGTCAAGGTCATTTGATTCGATAAATCAATCACCTGGGT 765
DB 661 AAGTTTAAAGATGAGAAATTTGGAGTCAAGTTTGGACTCGGTTAAATCAATCACCTGGGT 720
QY 766 GCTGAGGATGAGTTGAAGCATGAGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 825
DB 721 GCTGAGGATGAACTGAAGGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 780
QY 826 GTTGTCAAGGTTTCAATCTGTACAAAGAGGTTATACAGCAGTGCACACCTGTGCAAGA 885
DB 781 GTGATCACTGGTTTCAAGCTGTACAGAGCGGAGTTTACTAGCGACCATTTGTGAACT 840
QY 886 GATCCAATGATGTGAACCAAGCAGTTCTTGGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 945
DB 841 ACACCGATGATGTGAACCAAGCAGTTCTTGGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 900
QY 946 CCTTATTGGCTCATCAAGAACTCATGGGCTGACAAATTTGGGCTGACAAATGGCTACTTTAAG 1005
DB 901 CCTTATTGGCTCATCAAGAACTCATGGGCTGACAAATTTGGGCTGACAAATGGCTACTTTAAG 960
QY 1006 ATGGAATCGGCAAGAACTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1064
DB 961 ATGGAATCGGCAAGAACTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1019

D	b	827	TGCTTTTGTTCGGCCAGTTAGTGTCGCATTTGAGTAGCGAAGGACTTCCGATTCTACAA	886
Q	y	852	GAAAGGTGTATACAGCAGTGACACCTGTGGAAGAGATCCAATGATGTGAACCACGCCAGT	911
D	b	887	TAATGGAGTTTACACTAGTACCACTTTGTGTGAGCACGCCCATGGAATGTAATCATGCTGT	946
Q	y	912	TCTTGGCGTCGGTTATGGAATCGAGGACGGGATTCCTTATTTGGCTCATCAAGAATCATG	971
D	b	947	TCTTGTCTTGGGTATGAGTTGAAGATGGTGTTCATATTTGGATCATATAAAAATTCATG	1006
Q	y	972	GGGTACAAATTTGGGGTGACAATGCTACTTTAAGATGGAATCTCGGCAAGAACATGTGTGG	1031
D	b	1007	GGGAGCCACTGGGTGACAAATGGTTACTTTCAAGATGGAATTAGGGAGGAATATGTGCGG	1066
Q	y	1032	TGTTGCAACTTGGCATCTTATCCCATTTGGCT	1065
D	b	1067	TGTTGCAACTTGGCATCTTATCCCTGTGTGGCT	1100
 RESULT 11 US-10-425-114-18625 ; Sequence 18625, Application US/10425114 ; Publication No. US2004003488A1 ; GENERAL INFORMATION: ; APPLICANT: Liu, Jingdong ; APPLICANT: Zhou, Yihua ; APPLICANT: Kovalic, David K. ; APPLICANT: Screen, Steven E ; APPLICANT: Tabaska, Jack E ; APPLICANT: Cao, Yongwei ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement ; FILE REFERENCE: 38-21(53313)B ; CURRENT APPLICATION NUMBER: US/10/425,114 ; CURRENT FILING DATE: 2003-04-28 ; NUMBER OF SEQ ID NOS: 73128 ; SEQ ID NO 18625 ; LENGTH: 1390 ; TYPE: DNA ; ORGANISM: Glycine max ; FEATURE: ; OTHER INFORMATION: Clone ID: LTB3093-028-E3_FLI US-10-425-114-18625				
 Query March 45.7%; Score 489.6; DB 12; Length 1390; Best Local Similarity 69.1%; Pred. No. 7.6e-154; Indels 0; Gaps 0; Matches 669; Conservative				
Q	y	97	ACACAAAGGCGCTGACTGATGAGCCTGCCATCTCTCGCGCTTCGTCGAGTTGGCGGCAC	156
D	b	115	ATACGAATGTTGGCGGGGTGGAGCGAGGTGTTCCGTTGATAGGCACTGGCGGCGC	174
Q	y	157	GCCTTCCACTTCGACCGGTTCCCGCAGGTACGGAGAGACTACGGATCGAGAGGAG	216
D	b	175	CGCTGAAGTTTGTAGTTTATGAGCAGGTTCCGGAGAGATTACCGNAGCGAGGAAG	234
Q	y	217	ATCAAGAAGAGGTTCCGGATCTTCGTGGAGAACTAGCGTTTATCCCGTCCACTAATCGG	276
D	b	235	ATGAGGAGAGGTTACGAGATATTCTCGCAAACCTCAGGTTTCATCCGCTCCCAACAAG	294
Q	y	277	AAGATCTGTGATATCCCTAGGAATCAACCAATTCGCGGACCTGACCTGGGAGAAATC	336
D	b	295	AACCGTTGGCCCTACACTCTCTGTGTTATCATTTTTGCTGATTGCACTTGGAGGATTC	354
Q	y	337	CGGACCAATCGCCTTGGTGGCGCGACAACTGCTCGGCGATCGGCAATGGAACCCACGG	396
D	b	355	AAAGACACAGACTAGGCGCTGCCAAATTTGCTCTGCCACTTTAACGGCAACCAACAAG	414
Q	y	397	TTTGTGATGGCGTCTTCTGTAAACGAGGATTTGGAGGAGCAAGGATAGTGAGCCCT	456
D	b	415	CTCACTGATGCTGTTCTTCTCCCACGAAGACTGGAGGAAGAGAGGAATAGTGNATGAT	474
Q	y	457	GTAAGGACCAAGGAAGCTGTGATCTTGTGGACTTTTCAGTACTCTGGAGCACTAGAG	516

```
Db 475 GTTAAAGATCAGGCGAGCTGGGATCATGCTGGACATTCAGACAACTGGGCGTTAGAA 534
Qy 517 GCTGATATACACAGCTAACTGGAAGAGACATCATTTATCTGACAGCAACTTGTGGAC 576
Db 535 GCAGCCTATGCACAAGCAATTTGGGAAGAGTATCTCTCTTCTGAGCAGCACTAGTGGAT 594
Qy 577 TGTGCTCAGCAATTCATTAACCTTTGGATGCAATGAGGTTTGGCTTCCCAAGCCTTTGAA 636
Db 595 TGTGCTGCCGTTTCAACACTTTGGCTGTAATGGTGGTTGCCATCACAGCCTTTGAG 654
Qy 637 TACGTTAGTACAATGAGGAGCATGACACAGAAAGAGCTTATCCATACCTTGGTGTGAT 696
Db 655 TACATCAATACAAATGAGGAGCTTGAGACAGAGGAGCATACCCCTACACTGGAAAAGAT 714
Qy 697 GGTATCTGCAACTTCAAGCAGGAGATGTTGGTGTCAAGGTCATTTGATTCGATAAACATC 756
Db 715 GGTGATGCAAAATCTCAGCTGAAATGTTGCCGTTCAAGTCATTTGACTCGGTCAATATC 774
Qy 757 ACCCTGGTGTGAGAGTGAAGTGAAGCATGAGTGGCTTGGTGGTCCAGTTCAGGTT 816
Db 775 ACCTTGGGTGCTGAGAAATGAATTAACAATGCAATGCAATTTGTTGGCGCGTTAGTGTG 834
Qy 817 GCATTTGAGTGTGCAAGGTTTCAATCTGTACAGAAAGTGTATACAGCAGTGACACC 876
Db 835 GCCTTTCAGGTGGAATGGGTTCCATTTTACGAGAAATGAGTTTACACTAGTGACATT 894
Qy 877 TGTGGAAGAGATCCAAATGATGTGAACCAAGCAGTCTTTCGCGTGGTTATGAGTCCAG 936
Db 895 TCGCGTAGCACTTCCAGGATGTGAACCATGCGCTTCTGTTGGGATGAGGTTGAA 954
Qy 937 GACGGATTCCTTATGCTCATCAAGAACTCATGGGTGACAAATGGGTGCAATGGC 996
Db 955 AATGGCGTCCCATATTTGCTCATATAAAATTCATGGGAGAAAGTTGGGTTGGAATGGC 1014
Qy 997 TACTTTAAGATGGAATCGGCAAGAACATGTTGGTGTGTGCAACTTGGCACTTTATCCC 1056
Db 1015 TACTTCAAAATGGAATTTGGGAGAACATGTTGGTGTGTGCAACTTGTGGCTTTATCCA 1074
Qy 1057 ATTGTGC 1064
Db 1075 GTTGTGGC 1082
```

RESULT 12

```
US-10-424-599-122629
; Sequence 122629, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 122629
; LENGTH: 1933
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1933)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_81740C.1
US-10-424-599-122629
```

Query Match 45.0%; Score 481.8; DB 12; Length 1933;
Best Local Similarity 69.2%; Pred. No. 4e-151;

```
Matches 671; Conservative 0; Mismatches 297; Indels 1; Gaps 1;
Qy 97 ACACAAAGCGCTGACTCGATTGAGCTGCCATCCTCGGCGTCTTTCGACAGTTCGCCCCAC 156
Db 234 ATACGAATGTGTGCGGGGTGGAGCGGAGGTGTTCCGGTGTATCGGCAGATGCCCGCGC 293
Qy 157 GCCTTCCACTTCGACCGTTTCGCCCGCAGGTACGGGAAGAGCTTACCGATCGGAGGAGG 216
Db 294 GCGCTGAAGTTTCTAGGTTTCATGACGAGTTCGGGAAGAGTTTACCGAAGCGGAGGAG 353
Qy 217 ATCAAGAGAGGTTCCGGATCTTCGTTGGAGATCTAGCGTTTATCCGTTCCACTTAATCG 276
Db 354 ATGAGGAGAGGTACGAGATATTTCTCGAAACCTCAGGTTTCATCCGTTCCCAACAAG 413
Qy 277 AAGGATCTGTCTATACCTTAGGAATCAACCAATTCGCCGACCTGACCTGGGAGGAATTC 336
Db 414 AACCGCTTGCCCTTACACTCTCTCTGTTAATCATTTTGTGATTGGACTTGGGAGGAGTTC 473
Qy 337 CGACCAATCGCCTTGGTGGCGGAGAACTGCTCGGCACTCGCATCGGAACCAACCGG 396
Db 474 AAAAGACACAGACTAGCGCTGCCAAATTTGCTCTGCCACTTTAACGGCAACCAAG 533
Qy 397 TTTGTGATGCGCTGCTTCTGTAACGAGGAGTTGGAGGAGCAAGGAGTAGTGAGCCCT 456
Db 534 CTCACTGATGCTGTTCTTCTCCACGAAAGACTGGAGGAAGAAATAGTAGTAT 593
Qy 457 GTAAAGACCAAGAAAGCTGTGGATCTTGTCTGGAATTTTCAGTACTACTGAGCAGCTAG 516
Db 594 GTTAAAGATCAAGCGAGCTCGGATCATGCTGGACATTCAGCAACAATCGGGCTTTAGAA 653
Qy 517 GCTGCATATACACAGCTTAACGAGAGAGACATCATTTCTGAAACAGCAACTTGTGGAC 576
Db 654 GCACTTATGCAACAGCATTTGGAAAGATATCTCTCTTCTGAGCAGCAGCTAGTGGAT 713
Qy 577 TGTGCTCAGCACTTCAATAACTTTGGATGCAATGGAGTTTGCCTTCCCAAGCCTTTGAA 636
Db 714 TGTGCTGGCCTTTCAACAACTTTGGCTGTAAATGGTGGTGGTGCATCAACAAGCCTTTG 773
Qy 637 TAGC-TTAAGTACAATGGAGGATCGACACAGACAGACTTATCCATACCTTGGTCTCAA 695
Db 774 TACATTTAAATACAATGGTGGACTAGACACAGAGGAGCATATCCCTACACAGGAAAGA 833
Qy 696 TGGTATCTGCAACTTCAAGCAGGAGAAATGTTGGTGTCAAGGTCATTTGATTTCGATAAACA 755
Db 834 TGTGCTGCAAAATCTCAGCTGAAATGTTGCTGTTCAAGTCTCTTGAATAT 893
Qy 756 CACCTGGGTGCTGAGGATGAGTTGAAGCATGACAGTGGCGTTGGTCCGTTCCAGTTAGCG 815
Db 894 CACCTTGGGTGCTGAGAAATGAATTAACAATGCAATGCAATTTGTTCCGCGCTTTAGTGT 953
Qy 816 TGCATTTGAGGTTGTGAAAGGTTTCAATCTGTACAAAGAAAGGTGTATACAGCAGTGACAC 875
Db 954 GGCCTTTCAGTGTGTAATGGGTTCCATTTCTACGAGAAATGGAGTTTACACTAGTGACAT 1013
Qy 876 CTGTGAAAGAGATCCAAATGAGTGAACCAAGCAGTCTTTCGCGTGGTTATGAGTTCGA 935
Db 1014 TTGCGGTAGCACTTCCCAAGGATGTAAACCATGCGCTCTTGTGTGGGTTATGAGTTGA 1073
Qy 936 GGAAGGATTCCTTATTTGGCTCATCAAGAACTCATGCGGTACAAATTTGGGTGCACAATGG 995
Db 1074 GAATGGGCTCCCATATTTGGCTCATATAAAATTCATGGGAGAAAGTTGGGTGAGATGG 1133
Qy 996 CTACTTTAAGATGGAATCGGCAAGAACATGTTGGTGTGCAACTTGGCGCATCTTATCC 1055
Db 1134 CTACTTTCAAAATGGAATGGGAGAAACATGTTGGTGTGTTGCAACTTGTGCGCTTATCC 1193
Qy 1056 CATTGTGGC 1064
Db 1194 AGTTGTGGC 1202
```

RESULT 13

US-09-770-445-20/c


```
; Sequence 20, Application US/09770445
; Patent No. US20020023281A1
; GENERAL INFORMATION:
; APPLICANT: Goriach, Jern
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Kricker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2023US (PARA-012PRV)
; CURRENT APPLICATION NUMBER: US/09/770,445
; CURRENT FILING DATE: 2001-01-26
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 1282
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-770-445-20

Query Match 43.7%; Score 467.6; DB 9; Length 1282;
Best Local Similarity 66.7%; Pred. No. 1.9e-146;
Matches 668; Conservative 0; Mismatches 334; Indels 0; Gaps 0;

Qy 63 CGGTTTGAAGAGCAATCCATCGGTCGGTTACACAAAGCGCTGACTGATGAGCC 122
Db 1214 CGGATTCGATGATTAATCCATCAAAATGGTCTCCGACATCTCCATGAGCTTGAAGA 1155

Qy 123 TGCCATCTCGGCGTCTTGGCAGTTGCCGACGCTTCCACTCGCACGGTTGCCCG 182
Db 1154 CACGTCGTCAGATCTTAGACAGTCCCGTCAATGTTCTCTCTCTCTCTCTCTCTCT 1095

Qy 183 CAGGTACGGGAAGAGTACGATCGGATCGGAGGAGATCAAGAAGAGTTCGGATCTTCGT 242
Db 1094 CAGGTATGGGAAGAGTATCAGAGTGTGGAGGAGATCAAGATTCTCTGTTTTTAA 1035

Qy 243 GGAGATCTAGCGTTTATCGGTCCTAATCGGAAGATCTCGTATACCTAGGAAT 302
Db 1034 GGAGATCTTGAATTAATCAGATCCACTAATGAAGAGCTTATCTATAAATCTCTCT 975

Qy 303 CAACCAATTCGCGACTGACCTGGGAGGAATTCGGACCAATCGCCTTGGTGGCGCA 362
Db 974 TAATCAGTTTGTGCTGACTTGACATGGCAAGAGTTTCAAGATACAAGCTTGGAGCTCA 915

Qy 363 GAATGCTCGGACGTCGGATCGGAATCGAACCCCGTTTGTGATGGCGTCTCTCTTAC 422
Db 914 AAATGTTCTGCTACTTAAAGGTAGCCACGAGATCACTGAAGCTACAGTTCGAGACAC 855

Qy 423 GAGGATTTGAGGAGGAGCAAGGATAGTGAGCCCTGTAAAGGAGCAAGGAGCTGTGGATC 482
Db 854 AAAGATTTGAGAGAGATGTTATTTGTAGCCCTGTGAAGAACAGGACATTTGGATC 795

Qy 483 TTGCTGACATTCAGTACTGAGCACTAGAGCTGCATATACACAGCTAAGTGGAAA 542
Db 794 TTGTTGACATTTAGACAACTGGAGCTCTTGAAGCAGCTTACCATCAAGCATTTGGAAA 735
```

```
Qy 543 GAGCACATCATTAATCTGAACAGCAACTTGTGGACTGTGGCTCAGCATTTCAATTAACCTTGG 602
Db 734 AGGAATATCTTTGTCCGAGCAACAGCTTGTGGATTGTGTGGTACTTTCAATAACTTTGG 675

Qy 603 ATGCAATGGAGGTTTGGCTTCCCAAGCCTTTGAATAGTAACTCAATCAAGGAGCATCGA 662
Db 674 TTGTGATGGTGGACTTCTTCTCAAGCCTTTGAATACATTAATACACAGTGGGCTCGA 615

Qy 663 CACAGAACAGACTTATCCATACCTTGGTGTCAATGTGTATCTGCAACTTCAAGCAGGAGAA 722
Db 614 CACGAGGAGGCTTATCTTACACCCGAGAAAGCGGTGGCTGCAAAATTTTTCAGCGAGAAA 555

Qy 723 TGTGGTGTCAAGGTCAATTCGATTAACATCACCTCGGTGGCTGAGGATGAGTTGAA 782
Db 554 CATCGGTGTACAAGTCCGTGACTCTGTCAACATTTACCTCGGTGGAGAGATGACTGAA 495

Qy 783 GCATGCAAGTGGGCTTGGTGGCTCCAGTTCAGTTCAGTTTGAAGTGTGTGAAAGTTCAA 842
Db 494 GCAGCGGTCTGGTGGTAAAGCCAGTGTGGGCTTGAAGTGTGTATCAATGAATTCAG 435

Qy 843 TCTGTACAAGAAAGGTGTATACAGCAGTGCACCTGTGGAGAGATCCATGATGTGAA 902
Db 434 GTTTTATAGAAGGAGTGTTCCTAGCAATATATGTGTAAACACTCCAATGATGTAAA 375

Qy 903 CCACGCAATCTTGGCCGCTCGGTTATGAGTGCAGGACGGGATTCCTTATTTGGCTCATCA 962
Db 374 CCATGCAAGTGTAGCAGTTCGTTATGGAGTTGAAGACGAGCTCCCGTACTTGGCTTATA 315

Qy 963 GAATCATGGGGTACAAATTTGGGGTGCATAGTGGCTACTTTAAGATGGAATCGGCAAGAA 1022
Db 314 GAATCATGGGAGGTGAATTTGGGGGAGACAAATGGCTACTTCAAGATGGAATTTGGAAAGAA 255

Qy 1023 CATGTGTGGTGTTCGAACCTTGGCATCTTATCCATTTGTGGC 1064
Db 254 CATGTGTGGTGTTCGAACATGTTATCGTATCCGGTTGTAGC 213

RESULT 14
US-10-219-220-241
; Sequence 241, Application US/10219220
; Publication No. US20030082724A1
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; death and their use in the modification of plant development
; FILE REFERENCE: 11000.1022c1
; CURRENT APPLICATION NUMBER: US/10/219,220
; CURRENT FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: U.S. No. US20030082724A1 09/325,932
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 290
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 241
; LENGTH: 1652
; TYPE: DNA
; ORGANISM: Pinus radiata
; US-10-219-220-241

Query Match 38.0%; Score 407.4; DB 14; Length 1652;
Best Local Similarity 65.7%; Pred. No. 4.6e-126;
Matches 626; Conservative 0; Mismatches 321; Indels 6; Gaps 2;

Qy 116 TTGAGCTGCCATCTCTCGGCTCTTGGCAGTTGGCCGACGCTTCCACTTCGACGGT 175
Db 385 TGGAGTCCAAATTCGTTAAATCTTGGAAACCAATCCAAAGTCTCTCAATTTTCGGAGT 444

Qy 176 TCGCCCGCAGGTACGGGAAGAGCTTACGGATCGGAGGAGAGATCAAGAAGGTTTCGGGA 235
Db 445 TCGTGTGAGATAGGCAAGAGGTACGATCTGTCCATCAGCTTGTGCATAGATTCATG 504

Qy 236 TCTTCTGGAGAACTCTAGCGTTTATCCGGTCCACTAATCGAAGGATCTCTCGTATACCC 295
```


Db 505 TCTTGTGAAGAACTGGAGCTGATCGAGTCAAGAAACAGAAATGAAGCTTCTTACTT 564
Qy 296 TAGGAATCAACCAATTCGCCGACCTGACCTGGAGGAATTCGGACCAATCGCCTGGTG 355
Db 565 TGGCAATAAATGAGTTGCTGACATAACATGGAGGAATTCATGGACAAATTTGGGTG 624
Qy 356 CGGCGCAGAACTGCTGGCGAGCTGGCATGGAACACCGGTTTGTGATGGGTGCTC 415
Db 625 CTTACAGAACTGTTCCG---CTACCCACAGTAACCAATGACGTATGCCAGCTTC 681
Qy 416 CTGTAAACAGGGAATTTGGAGGAGCAAGGATAGTGCCTGTAAAGACCAAGGAAGCT 475
Db 682 CTGCGAAGAAAGACTGGAGACAGAAGGACATAGTGAGTCTCTGTAAGAAACCAAGCCCAT 741
Qy 476 GTGGATCTGTGGACTTTTCACTACTGAGCACTAGAGCTGCATATACAGACTAA 535
Db 742 GTGGATCTGTGGACATTCAGCACTACTGAGCACTAGAGCTGCCTATCTCAGGCTA 801
Qy 536 CTGGAAGAGACACATCATATCTGAACAGCACTTGTGCACTGTGCCTCAGCATTCATA 595
Db 802 CAGGAAGACATGTTATCTGTCTGAACAGCAGCTGTTGACTGTGCTGGACCAATTAACA 861
Qy 596 ACTTTGGATGCAATGGAGGTTTCCCTCCAGGCTTTTGAATAGCTTTAAGTACAAATGGAG 655
Db 862 ACTTTGGTTCAAATGGTGAAGCTGCCATCCCAAGCAATTTGAGTATATCAAGTACAAACGGAG 921
Qy 656 GCATCGACACAGACAGACTTATCCATACCTTGTGTCATGTTGTCATGTTGCAACTTCAAGC 715
Db 922 GCTTGACACTGAGGAAGCCTATCTTACAGGCTAAGATGGTGTGTTGTAATATGATG 981
Qy 716 AGGAATATGTTGGTTCAGGTCATGATGATTCGATAAACAATCAACCCCTGGTGTGAGGATG 775
Db 982 TAAACAATGTCGGTGAAGGTTGCTGATAGTGTCAACATCAGATTTGGGGGCAGAAGATG 1041
Qy 776 AGTGAAGCATGCACTGGGCTTGTGCGTCCAGTTAGGTTGCTGATTTGAGGTTCTGAAAG 835
Db 1042 ACTCAAGCTGCTGTGGGTCTAGTTCGCCCTGTGAGTGGCATTCGAAGTTAATTCGG 1101
Qy 836 GTTTCATCTGTACAAAGAGTGTATACAGCAGTGACACCTGTGGAAGAGATCCAATGG 895
Db 1102 AATTTGATTTTACAAAGAGAGTCTTTTACAGCACAAGTTGTGTTCAAGGCCCAATGG 1161
Qy 896 ATGTGAACACGACGTTCTTCCGCTGGTATGAGT---CGAGGACGGGATTCCTTAT 952
Db 1162 ATGTCAACCATGCTCTTTTGGCCGTGGGTATGTTGTTAGTGAGGAGGGGACTCCACAT 1221
Qy 953 GCTCATCAAGAACTCATGGGGTACAAATTTGGGGTGACAATGGGCTACTTTAAGATGGAAC 1012
Db 1222 GATCATCAAGATTCCTGGGAAGAGAGCTGGGTGTTGATGGATCTTCAAGATGGAGT 1281
Qy 1013 TCGGCAAGAACATGTGTGGTGTGCAACTTGGCAATCTTATCCCATTTGGCT 1065
Db 1282 TAGGAAGAATATGTGTGTGTGGACTTGTGCTTGTACCCCTATTGTTTCT 1334

RESULT 15

US-10-425-114-3201
; Sequence 3201, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Vihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 3201

; LENGTH: 830
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700243475_FLI
US-10-425-114-3201

Query Match 34.5%; Score 369; DB 12; Length 830;
Best Local Similarity 77.5%; Pred. No. 2.8e-113; Indels 0; Gaps 0;
Matches 447; Conservative 0; Mismatches 130;
Qy 488 GGAATTTTCACTACTTGGAGCACTAGAGCTGCATATACAGCTTAATCTGGAATGCA 547
Db 2 GGACCTTCAGCACTACTTGTGCACTTGAAGCTGCATATACCCAGGCACTGGCAAGCCCA 61
Qy 548 CATCATTTCTGAACGAACCTTGTGGACTGCTCAGCATTCATTAATCTTGGATGCA 607
Db 62 TCTCTCTCTGTGAGCAACAGCTTATTGACTGTGGTGTTCATTTCAACAATTTCCGATGCA 121
Qy 608 ATGGAGGTTTGCCTTCCCAAGCCTTTGAATACCTTAAGTACATGAGGAGCATCGACACAG 667
Db 122 ACGAGGCTTCCATCCAGGCTTTGAATACATCAATACATGATGGCTTGCACATG 181
Qy 668 AACAGACTTATCCATACCTTGGTGTCAATGGTATCTGCAACTTCAAGCAGGAGAAATGTTG 727
Db 182 AGGAATCTTACCTTACCAAGGTGTCAATGGAATCTGCAAGTTTAAAGAAATGAGATGTTG 241
Qy 728 GTCTCAAGGTTCATTGATTCGATAAACAATCACCTGGGTGCTGAGGATGAGTTGAAGCATG 787
Db 242 GATCAAGGTTTGGACTGGTTAATCACCTGGGTGCTGAGGATGAGTTGAAGCATG 301
Qy 788 CAGTGGGCTTGGTGCCTCCAGTTAGCTTGCATTTGAGGTTGTGAAAGGTTTCAATCTGT 847
Db 302 CTGTTGCTGTGGTTCGCCAGTTAGTGTTCCTTCGAGGTGATCACTGGTTTCAGGCTGT 361
Qy 848 ACGAAGAGGTGTATACAGCAGTGACACCTGTGGAAGAGATCCAATGGATGTGAACACG 907
Db 362 ACAAGAGCGAGTTTACACTAGCGACCTTGTGGAACCTACCGATGGATGTGAACACG 421
Qy 908 CAGTCTTTCGCGTCCGTTATGAGTCGAGACGGGATTCCTTATTCGCTCATCAAGAACT 967
Db 422 CTGTTCTGCTGTTCGCTACGCTGTCGAAGATGGTGTACCTACTGCTCATCAAGAACT 481
Qy 968 CATGGGGTACAAATTTGGGTGACAATGGCTACTTTAAGATGGAACCTCGGCAAGAACTGT 1027
Db 482 CATGGGGCGTGAATGGGGTGTAGGGGTACTTTCAAGATGGAATGGGCAAGAACTGT 541
Qy 1028 GTGCTGTTCGAACCTTCGCGCATCTTATCCCATTTGTGGC 1064
Db 542 GCGGTGTTGCTACGTGTGCATCTTATTTGTGCGC 578

Search completed: March 29, 2004, 22:09:43

Job time : 332 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: March 29, 2004, 19:15:42 ; Search time 2391 Seconds
(without alignments)
13376.159 Million cell updates/sec

Title: US-10-087-714-1
Perfect score: 1071
Sequence: 1 atggcagctaaagctctctt.....atccattgtggtgtgtag 1071

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estnu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_estrc.*
9: gb_est1.*
10: gb_est2.*
11: gb_est3.*
12: gb_est4.*
13: gb_est5.*
14: gb_est6.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_pbg.*
27: em_gss_vrl.*
28: gb_gss1.*
29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	571.2	53.3	1532	11	AV106223
2	434.6	40.6	989	14	CK263045
3	430.8	40.2	829	14	CB341855
4	417.8	39.0	1005	14	CF212812

5	410.8	38.4	814	12	BG594735
6	410	38.3	985	14	CK287181
7	408	38.1	731	13	BQ281381
8	407.8	38.1	748	14	CB982399
9	405.6	37.9	936	14	CK268548
10	403.4	37.7	920	14	CK257583
11	402.8	37.6	848	14	CF510505
12	401.8	37.5	886	14	CF510584
13	398.6	37.2	977	14	CK295860
14	394.2	36.8	720	10	BE131652
15	390.8	36.5	965	14	CK294622
16	389.8	36.4	910	13	BQ797116
17	387.8	36.2	794	10	BE195255
18	386.2	36.1	768	14	CB892658
19	384.8	35.9	701	14	CB894684
20	383.8	35.8	653	14	CB917819
21	383.4	35.8	924	14	CK294694
22	378.6	35.4	755	14	CB981790
23	378.6	35.4	930	14	CK295057
24	378.4	35.3	1230	10	BE422216
25	378	35.3	709	12	BG155091
26	377	35.2	852	14	CK263154
27	375.2	35.0	639	10	AW155820
28	375.2	35.0	665	10	AW399964
29	375	35.0	863	29	CG789556
30	373.6	34.9	760	10	BF267048
31	372.6	34.8	894	13	BU039998
32	370.8	34.6	609	14	CB006195
33	370.6	34.6	827	9	AA979924
34	369.6	34.5	901	14	CK298522
35	368.4	34.4	635	13	BU039803
36	367.8	34.3	946	14	CK285936
37	367.4	34.3	710	12	BI267420
38	366.6	34.2	703	12	BI269594
39	366.2	34.2	815	14	CA809544
40	365.4	34.1	648	13	BQ625193
41	365.2	34.1	888	14	CF519040
42	364.8	34.1	911	14	CK286288
43	364.4	34.0	888	14	CK292918
44	363.8	34.0	755	12	BG45271
45	362.6	33.9	833	14	CB666317

ALIGNMENTS

RESULT 1	AY106223	1532 bp	mRNA	linear	HTC 16-OCT-2002
LOCUS	AY106223	Zea mays	PC0114759	mrna	sequence.
DEFINITION	AY106223	Zea mays	PC0114759	mrna	sequence.
ACCESSION	AY106223	HTC			
VERSION	AY106223.1	GI:21209301			
KEYWORDS	HTC				
SOURCE	Zea mays				
ORGANISM	Zea mays				
REFERENCE	1	(bases 1 to 1532)			
AUTHORS	Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S., Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.				
TITLE	Maize Mapping Project/Dupont Consensus Sequences for Design of Overgo Probes				
JOURNAL	Unpublished (2002)				
REFERENCE	2	(bases 1 to 1532)			
AUTHORS	Coe, E.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA				
COMMENT	If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSU, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the				

1	571.2	53.3	1532	11	AV106223
2	434.6	40.6	989	14	CK263045
3	430.8	40.2	829	14	CB341855
4	417.8	39.0	1005	14	CF212812

2d. Set 4 were grown under the standard conditions and then were heat stressed by placement at 35 C. Heat stressed leaves were harvested at 2 hr, 6 hr, 12 hr, 1 d, 2d and 4d and heat-stressed roots were harvested at 6 hr, 12 hr, 1 d, and 4d. RNA was isolated from all tissues and equal RNA from each tissue and stress was pooled to construct the cDNA library. RNA sample."

ORIGIN

Query Match 40.6%; Score 434.6; DB 14; Length 989;
 Best Local Similarity 68.5%; Pred. No. 2.9e-94;
 Matches 599; Conservative 0; Mismatches 276; Indels 0; Gaps 0;

QY 107 CTGACTCGATTGAGCTGCCATCTTCGGCTCGCGTCTTGGCAGTTGCCGCCACGCCCTTCCACT 166
 DB 115 CCGAGAGCTGGAGACGGAACTCTTCAAGTCTCGCCAGACTCGCAATGCTCTCTCT 174

QY 167 TCGACGGTTCCCGCGAGTACGGAGAGAGCTACGGATCGGAGGAGAGATCAAGAGA 226
 DB 175 TCGCTCGCTTTGCTATCAGGATCGGAAAGGTACAGTCCGTTGAGGAGATCAAGCAA 234

QY 227 GGTTCGGGATCTTCGTGGAGATCTAGCGTTTATCCGCTCCACTAATCGGAGGATCTGT 286
 DB 235 GGTTCGAGATATTTTGGCAATCTGAAGATGATCCGATCGCATACAGCAAGGACTAT 294

QY 287 CGTATACCTAGGAATCAACCAATTCGCCGACCTGACCTGGAGGAAATCCGGACCAATC 346
 DB 295 CATACAAATCTCGTGTCAATGAGTTTACCGACCTAATGATGGATGAGTTCGCTAGACACA 354

QY 347 GCCTTGGTGGGCGAGACTCTCGGCGACTGCGCATGCGCATGGAAACACCGCTTTCGATG 406
 DB 355 AGTTGGGGCATCTCAAACTGTTCTGCCACTACAAAGGGCAATCTCAAGCTAATTAACG 414

QY 407 GGGTGTCTTCTTAACGAGGATTTGGAGGAGCAAGGGATAGTGAGCCCTGTAAAGGACC 466
 DB 415 TTGTTCTGCGACAGAGGAGACTGGAGGAGATGGTATTGTTAGCCAGTGAAGGCAC 474

QY 467 AGGAAGCTGTGGATCTTCTGACTTTCAGTACTACTGGACACTAGAGCTGATATA 526
 DB 475 AGGGCAAGTGGCGGATCTTCTGCGACATTCAGCACTACTGTTGCTACTAGGAGCAGATG 534

QY 527 CACAGTAACTGGAAGACACATCATTTATCTGAACAGCAACTTGTGACTGTGCTCAG 586
 DB 535 CCGAGCATTTGGAGGAACTCTCTGTCAGAGCAGCAGCTTGTGACTGTGCTGGAG 594

QY 587 CATTAATTAATTTGATGCAATGAGGTTTGGCTTCCCAAGCCTTTGAATACGTTAAGT 646
 DB 595 CTTTAAATAATTTGGCTGCAATGGGGGTTGCCATCAAGCCCTTTGAGTACATTAAT 654

QY 647 ACAATGAGGCATCGACACAGACACTTATCCATACCTTGGTGTCAATGTATCTGCA 706
 DB 655 TCAATGGTGTCTTGACACTGAAGAGATATCCATACACCGGCAAGATGGCATATGTA 714

QY 707 ACTTCAAGCAGGAGATGTTGGTGTCAAGGTCATTTGATTGATTAACATCACCTTGGGTG 766
 DB 715 AATTCTCACAGCAATATTGGTGTCAAAGTCATCAGTTCTGTCAATATTACCTGGGTG 774

QY 767 CTGAGATGATGTTGAAGCATGAGTGGGCTTGGTGGCTCCAGTTAGCTTGCATTGAGG 826
 DB 775 CTGAAGATGAATGAATACGAGTTCGATTTGGTGGCTGTTAGTGTGCTTTTGGAG 834

QY 827 TTGTGAAGGTTTCAATCTGTACAGAAAGGTGTATACAGCAGTGCACACTCTGGAAGAG 886
 DB 835 TGGTAAGGTTTCAACAGATATAGAGCGGAGTTTACACAGCACTGAATGTGCGGACA 894

QY 887 ATCCATGAGATGTAACCGACGAGTTCTTGGCGTGTGGTTATGAGTTCGAGGCGGATTC 946
 DB 895 CTCCCATGGACGTAACCACTGCTGTTCTTGGTGGGGTACGGTGTGTAATAATGCGGTC 954

QY 947 CTTATTGGCTCATCAAGAACTCATGGGTCACAAAT 981
 DB 955 CCTACTGGCTCATAAAGAACTCATGGGAGCAGAT 989

RESULT 3

CB341855 829 bp mRNA linear EST 10-APR-2003
 CA32EN0002_IIIBF_H04 Cabernet Sauvignon Leaf - CA32EN Vitis
 vinifera cDNA clone CA32EN0002_IIIBF_H04 5', mRNA sequence.

ACCESSION

CB341855

VERSION

CB341855.2 GI:29785118

KEYWORDS

EST.

SOURCE

Vitis vinifera

ORGANISM

Vitis vinifera
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; Vitaceae; Vitis.

REFERENCE

1 (bases 1 to 829)

AUTHORS

Goes da Silva, F., Lim, H., Iandolino, A., Baek, J., Jones, K.,
 Walker, M.A. and Cook, D.R.

TITLE

Transcriptional responses of Vitis vinifera to infection by the
 bacterial pathogen Xylella fastidiosa

JOURNAL

Unpublished (2003)

COMMENT

On Mar 14, 2003 this sequence version replaced gi:28962822.
 Contact: Douglas Cook, PhD
 CAES Genome Facility

FEATURES

UC Davis, Plant Pathology
 One Shields Ave, Davis, CA 95616, USA
 Tel: 530 754 6561
 Fax: 530 754 6617
 Email: drcook@ucdavis.edu

source

Seq primer: ACGTACCGACATATGCC.

Location/Qualifiers

1..829
 /organism="Vitis vinifera"
 /mol_type="mRNA"
 /cultivar="Cabernet Sauvignon"
 /db_xref="taxon:29760"
 /clone="CA32EN0002_IIIBF_H04"
 /sex="Hermaphrodite"
 /dev_stage="Mid-season leaf material, collected July 25, 2001."
 /lab_host="DH5alpha"
 /clone_lib="Cabernet Sauvignon Leaf - CA32EN"
 /notes="Organ: Leaf; Vector: pDNR; Site: 1: SfiI; Site 2: SfiI; CA32EN is a cDNA library of Cabernet Sauvignon leaves. The leaves were collected on July 25, 2001, in Napa Valley, California, and represent leaves in mid-season development. These leaves were asymptomatic and verified to be non-infected with the bacterial pathogen, Xylella fastidiosa, based on a diagnostic assay using PCR and Xylella-specific primer pairs. cDNA made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:
 5'-AAGCAGTGGTATCAACGACAGTGGCCATTACGCGCGG-3' and
 5'-ATTCTAGAGCGCGGCGGACATG-dt(30)NN-3'. Library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5-3 kb size fraction."

ORIGIN

Query Match 40.2%; Score 430.8; DB 14; Length 829;
 Best Local Similarity 70.7%; Pred. No. 2.2e-93;
 Matches 573; Conservative 0; Mismatches 237; Indels 0; Gaps 0;

QY 125 CCATCTCTCGCGCTCTTGGCAGTTGCCGCCACGCTTCCACTTCGACAGTTCGCCCGCA 184
 DB 3 CCGTCTCTCGCGCTGATCGGGACACGCGTCACTCTTCTCGCGAGCTTCGCTCACA 62

QY 185 GGTACGGGAGAGCTACGGATCGGAGGAGAGATCAAGAGAGTTCGGGATCTTCCTGG 244
 DB 63 GGTATGGAGAGAGTTCAAGCGGTGGACGAGATTAAGCTGAGATTCAGATTTTCGG 122

QY 245 AGAATCTAGCGTTTATCCGGTCCACTAATCGGAGGATCTGTGATATACCTAGGAATCA 304
 DB 123 AGAATTTGAAACTCATCAGATCCACCAACAGAAAGGCTTGCCTTATCTACTAGCTGTA 182

```
QY 305 ACCAATTGCGGACCTGACCTGGAGGAATTCGGACCAATCGCCTTGGTGGCGGCGAGA 364
Db 183 ATCAGTTCCGCTGATTGGACCTGGGAAGATTCGACAGACACAGGTTGGGAGCTGCTCAGA 242
QY 365 ACTGCTGGGACCTGGCATGGAACACACCGGTTTTCGATGGCGTCTCTCTCTGTAACGA 424
Db 243 ACTGCTCTGCCACCTTGAAGGCAATCACAGCTAACTGACGTTATCTCTCTGAGACGA 302
QY 425 GGGATTGGAGGAGCAAGGATAGTAGCCCTGTAAAGGACCAAGGAAGCTGTGGATCTT 484
Db 303 AAGATTGGAGAGAGATGGCATAGTCAGCCCAATCAAAGATCAAGGTCACTGTGGATCTT 362
QY 485 GCTGGACTTTCAGTACTAGGACACTAGGAGCTGATAGGCTGATATACACACTAACTGGAAGA 544
Db 363 GCTGGACTTTCAGACCACTGGAGCTCTAGAGGAGCTTACGCTCAGGCAATTTGGGAGG 422
QY 545 GCACATCAATTTCTGAACAGCAACTTTGGTGAAGTGTGCTCAGCAATTCATTAACCTTTGGAT 604
Db 423 GGATCTCTCTGCTGAGCAGCAGCTTTGGGACTGTGCGGAGCTTTCAATAACTTTGGAT 482
QY 605 GCAATGGAGGTTTGGCTTCCCAAGCCTTTGAATACGTTAGTACAAATGGAGGACATGACA 664
Db 483 GCCACGGGGGATGCCATCCCAAGCTTTTGGATGATCAATCAATCAATGGTGGCTTGATA 542
QY 665 CAGAACAGACTTATCCATACCTTGGTGTCAATGGTATCTGCAACTTCAAGCAGGAGATG 724
Db 543 CTGAGGAGCATATCTTACACTGGACTAGATGGCACCCTGCAATTTTCTTCAGAAAATA 602
QY 725 TTGGTGTCAAGGTCAATGATTCGATAAATCATCACCTGGTGGTCTGAGGATGAGTTGAGC 784
Db 603 TTGGTGTTCAGATTTCTGCACTCTGTGAATATTTACCCCTGGTGGTGAAGATGAGTTAAGC 662
QY 785 ATGAGTGGCTTGGTGGTCCAGTTAGCTTGGATTTGAGTTGTGAAGGTTTCAATC 844
Db 663 ATGCAAGTTCGATTTGCTGTCAGTACGTTGGATTTGAGTGGTCCATGATTTCCGAT 722
QY 845 TGTACAGAAAGGTGTATACAGCAGTGACACCTGTGGAAGAGATCCCAATGATGTGAACC 904
Db 723 TTTACAGAAAGGAGTTTACACAGTGGAACTTTGTGACAGCAGCTCCCATGATGTGAACC 782
QY 905 ACGAGTCTTGGCTGGTGGTATGGAGTCG 934
Db 783 ATGCTGTTCTTGCAGTCGGGTATGGAGTTG 812
```

```
RESULT 4
CF212812
LOCUS
DEFINITION
  CGF100651.H04 Vitis vinifera cv. cabernet sauvignon Stem - CAST
  Vitis vinifera cDNA clone CAST0005_IF_H04 5', mRNA sequence.
ACCESSION
  CF212812
VERSION
  CF212812.1 GI:33407185
KEYWORDS
  EST.
SOURCE
  Vitis vinifera
  Vitis vinifera
  Vitis vinifera
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  rosids; Vitaceae; Vitis.
  1 (bases 1 to 1005)
  Goes da Silva, F., Iandolino, A., Lim, H., Baek, J., Jones, K. and
  Cook, D.
  Expressed sequence tags from Vitis vinifera 'Cabernet sauvignon'
  berries at various developmental stages
  Unpublished (2003)
  Contact: Douglas Cook, PhD
  CAES Genome Facility
  UC Davis, Plant Pathology
  One Shields Ave, Davis, CA 95616, USA
  Tel: 530 754 8561
  Fax: 530 754 8617
  Email: drcook@ucdavis.edu
  Seq primer: ACGGTACCGACATATGCC.
  Location/Qualifiers
```

FEATURES

source

```
1..1005
/organism="Vitis vinifera"
/mol_type="mRNA"
/cultivar="Cabernet Sauvignon"
/db_xref="taxon:29760"
/clone="CAST0005_IF_H04"
/sex="Hermaphrodite"
/dev_stage="Pre-bloom (10-11 days before bloom)"
/lab_host="DH5alpha"
/clone_lib="Vitis vinifera cv. cabernet sauvignon Stem -
CAST"
/note="Organ: Stem; Vector: pDNR; Site 1: sf11; Site 2:
sf11; CAST is a cDNA library of Vitis vinifera cv.
'Cabernet Sauvignon' stems. Samples were collected May 13,
2002 from pre-bloom plants (10-11 days before bloom),
pre-veraison. Sampled vines were located at the University
of California, Davis, Experimental Vineyard. cDNAs were
made by oligo-dT priming and directionally cloned. 5' and
3' adaptors were used in cloning as follows:
5'-AAGCAGTGTATCAACGAGTGGCATTACGCGCGG-3' and
5'-ATTCTAGAGCGGCGGCGACATG-dt(30)NN-3'. Library was
constructed using the Clontech Creator SMART kit and
size-selected to contain the 0.5-3 kb size fraction."
```

ORIGIN

```
Query Match 39.0%; Score 417.8; DB 14; Length 1005;
Best Local Similarity 68.6%; Pred. No. 3.4e-90;
Matches 589; Conservative 0; Mismatches 268; Indels 1; Gaps 1;

QY 65 GTTTCGAGAGACAAATCCAAATCCGGTCCGTTACACAAAGGCTGACTCGATTGAGCCTG 124
Db 148 GCTTCGATGAGAGAACCAATATAGATGGTATTCGAGCAGCATACGCGACTTGGATCGT 207
QY 125 CCATCTCTCGGCTCTTGGCAGTTGCGCGACGCTTCCACTTCGACGCTTCCGCGCGCA 184
Db 208 CGTCTCTCGGCTGATCGGGGACACGCTCACGCTCTCTCGGAGCTTCGCTCACA 267
QY 185 GGTACGGGAGAGCTACGGATCGGAGGAGAGATCAAGAGAGGTTCCGAGTCTCGTGG 244
Db 268 GGTATGGGAGAGTTTACAGAGCGTGGACGAGATTAAGCTGAGATTCGAGATTTTCTCG 327
QY 245 AGAATCTAGGTTTATCCGGTCCACTTAATCGGAAGATCTGTCGTATACCTAGGAATCA 304
Db 328 AGAATTTGAACATCATCAGATCCACCACAGAAAGGCTTGCTTATCTAGCTGTTA 387
QY 305 ACCAATTCGCGACCTGACCTGGAGGAATTCGGACCAATCGCCTTGGTGGCGGCGAGA 364
Db 388 ATCAGTTCTGCTGATTGGACCTGGGAAGAGTTCCGACAGACACAGGTTGGAGCTGCTCAGA 447
QY 365 ACTGCTCGGCGACTGGCATGGAAACACCGGTTTGTGATGGCGTCTCTCTGTAACGA 424
Db 448 ACTGCTCTGCCACCTTGAAGGCAATCAACAGCTAACTGACGTTATCTCTCTGAGACGA 507
QY 425 GGGATTGGAGGAGCAAGGATAGTAGCCCTGTAAAGGACCAAGGAAGCTGTGGATCTT 484
Db 508 AAGATTGGAGAGAGATGGCATAGTCAGCCCAATCAAAGATCAAGGTCACTGTGGATCTT 567
QY 485 GCTGGACTTTCAGTACTACTGAGGACATAGAGCTGCATATACACAGCTAACTGGAAGA 544
Db 568 GCTGGACTTTCAGCACCCTCTAGAGGAGCTTACGCTCAGGCAATTTGGGAGG 627
QY 545 GCACATCAATTTCTGAACAGCAACTTTGGGAGCTGTGCGCTCAGCATTTCAATAACTTTGGAT 604
Db 628 GGATCTCTCTGCTGAGCAGCAGCTTTGGGACTGTGCGGAGCTTTCAATAACTTTGGAT 687
QY 605 GCAATGGAGGTTTGGCTTCCCAAGCCTTTGAATACGTTAGTACAAATGGAGGACATGACA 664
Db 688 GCCACGGGGGATGGCATCCCAAGCTTTTGGATGATCAATCAATGATGGCTTGATA 747
QY 665 CAGAACAGACTTATCCATACCTTGGTGTCAATGGTATCTGCAACTTCAAGCAGGAGATG 724
Db 748 CTGAGGAGCATATCTTACACTGGACTAGATGGCACCCTGCAATTTTCTTCAGAAAATA 807
```

```

QY 725 TTGGTCTCAAGTTCATTGATTCGATAAATCAATCACCTCGGTGCTGAGGATGATTTGAAGC 784
Db 808 TTGGTCTCAAGTTCCTCGACTCTGTGAATATTACCTCGGTGCTGAGGATGATTTAAAGC 867
QY 785 ATGCACTGGGCTTTGGTGGCTGCAGTTAGCGTTGTCATTTCAGGTTGTGAAAGGTTTCAATC 844
Db 868 ATGCACTGGCATTGTTGTTGCTCCAGTGAAGTGTGCAATTTTCAGGTGGTCCATGATTTCCGAT 927
QY 845 TGTACAGAAAGTGTATACAGCAGT-GACACCTGTGGAGAGATCCCAATGATGTGAAC 903
Db 928 TTACGATCAGGAGTTTACAAAGTGGAACTTTGGCAGCACTCCCAATGATGTGAAC 987
QY 904 CACGCACTTCTTGGCCGTC 921
Db 988 CATGCTGTTCTTGCACTC 1005

RESULT 5
BG594735 814 bp mRNA linear EST 07-MAR-2003
DEFINITION EST493413 cSTS Solanum tuberosum cDNA clone CSTS8E24 5' sequence,
mRNA sequence.
ACCESSION BG594735
VERSION BG594735.1 GI:13612875
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum.
REFERENCE van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Chiemingo,A.,
Bougri,O., Buell,C.R., Roming,C., Tanksley,S. and Baker,B.
Generations of ESTs from sprouting potato eyes
Unpublished (2000)
CONTACT: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
This clone can be obtained from the University of Arizona Genomics
Institute. Orders can be made through URL:
http://genome.arizona.edu/orders/
Seq primer: M3F-R.

FEATURES
Location/Qualifiers
1..814
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="CSTS8E24"
/tissue_type="sprouting eyes from tubers"
/dev_stage="12-14 weeks post harvest"
/lab_host="SOLR"
/clone_lib="cSTS"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; Various sizes of sprouting eyes (2mm to 15mm) were
taken from tubers. The tubers were incubated at 26C in the
dark for 2-3 weeks prior to sprouting. The eyes were
frozen in liquid nitrogen immediately upon removal from
tubers."

ORIGIN
Query Match 38.4%; Score 410.8; DB 12; Length 814;
Best Local Similarity 70.0%; Pred. No. 1.5e-88;
Matches 567; Conservative 0; Mismatches 242; Indels 1; Gaps 1;

QY 165 CTTGCAACGGTTCGGCCGAGTACGGGAAGAGTACGGATCGGAGGAGAGATCAAGAA 224
Db 6 CTTGCGTCGCTTGTGATCAGCATCGGAAGGTAACGATCCGTTGAGGAGATCAAGCA 65
QY 225 GAGGTTCCGAGATCTTCGTGGAGAACTACGGTTTATCCGGTCCCACTAATCGGAAGGATCT 284
Db 66 AAGGTTTCGAGATATTTTGGCAATCTGAAGATGATCCGATCGCATACAGCAAGGACT 125

```

```

QY 285 GTCTATACCTAGGAATCAACCAATTCGCGACCTGACCTGGAGGAAATTCGGACCAA 344
Db 126 ATCATACAAATCTCGTGTCTCAATGATTTACCGCACTTAACATGGGATGAGTTCCGTAGACA 185
QY 345 TCGCTTTGGTGGCGCGAGAACTGCTCGGACGCTGCGCATGGCAACCAACCGTTTTCGA 404
Db 186 CAACTTGGGGGCACTCTCAAACTGTTCTGCCATCAAAAGGGCAATCTCAAGCTAACTAA 245
QY 405 TGGCGTCTTCTCTGTAAACGAGGGANTGGAGGGAGCAAGGATAGTGAGCCCTGTAAAGGA 464
Db 246 CGTCTTCTGCCAGAGCAAGGACTGAGAGGAAAGTGGTATTTGTTAGCCCACTGAAGGC 305
QY 465 CCAGGAAGCTGTGGATCTTCTGGACTTTTCAGTACTACTGGAGCACTAGAGGCTGCATA 524
Db 306 ACAGGCAAGTGGGATCTTCTGGCAATTCAGCACTACTGGTGCTAGAGGAGGAGATA 365
QY 525 TACACAGCTAACTGGAAAGAGCACATCATTTATCTGAACAGCAACTTGTGGACTGTGCTC 584
Db 366 TGCCCAAGCACTTTGGGAAGGGAATCTCTCTGTGAGAGCAGCAGCTTGTGGACTGTGCTGG 425
QY 585 AGCAATTCAAATACTTTGGATGCAATGGAGGTTTGGCTTCCCAAGCCTTTGAATACGTAA 644
Db 426 AGCTTTTAAATACTTTGGCTGCAATGGGGGTTGCCATCAAAAGCCTTTGAGTACATTA 485
QY 645 GTACATGGAGGATCGACACAGAACTATTCATACCTTGGTGTCTCAATGATATCTG 704
Db 486 ATTCAATGGTGTCTTGACACTGAAGAGCATATCCATACACGGCAAGATGGCATATG 545
QY 705 CAATTCACAGAGAGAAATCTTGGTGTCAAGGTCATTGATTCGATAAACATCAACCTGGG 764
Db 546 TAAATTTCTCAAGCAAAATATTGGTGTCAAAGTCATCAGTTCTGTCAATATTACCTGGG 605
QY 765 TGCTGAGGAGTGTGAAGCATGCAGTGGCTTGGTCCGTCAGTTAGCGTTGCTGATTTGA 824
Db 606 TGCTGAAGATGAATGAAATACGAGTTGCAATTTGGTTAGGCCCTGTAGTGTGCTTTGA 665
QY 825 GGTGTGAAAGGTTTCAATCTGTCAAGAAAGGTGTATACAGAGTGACACCTGTGGAAG 884
Db 666 GGTGTGAAAGGTTTCAACAGTATTAAGAGCGAGTTTACACAGCACTCAATGTGGCGA 725
QY 885 AGATCCCAATGATGTGAACCAAGCAGTTCTTGGCGTGGTTATGGAGTCGAGGACGGAT 944
Db 726 CACTCCCATGGAGTAAACCATCTGTCTTGTGTGGGCTCGGTGGTGAATAATGCGCT 785
QY 945 TCCTTATTGGCTCATCAAGAACTCATGGGG 974
Db 786 TCCTTACTGCTCAT-AAGAACTCATGGGG 814

RESULT 6
CK287181 985 bp mRNA linear EST 15-DEC-2003
LOCUS CK287181
DEFINITION EST749903 Nicotiana benthamiana mixed tissue cDNA library,
normalized, full-length Nicotiana benthamiana cDNA clone NEMB738 5'
end, mRNA sequence.
ACCESSION CK287181
VERSION CK287181.1 GI:39863480
KEYWORDS EST.
SOURCE Nicotiana benthamiana
ORGANISM Nicotiana benthamiana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Nicotiana.
REFERENCE Buell,C.R., Hart,A., Zismann,V., Karamycheva,S.A., Day,B.,
Staskiewicz,B., Jin,H. and Baker,B.
Generation of EST sequences from Nicotiana benthamiana
Unpublished (2003)
Other ESTs: EST749904
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA

```


TJ Close lab (Choi, Close, Fenton) at the University of California, Riverside. The cDNA clones were in vivo excised to give pBluescript phagemids before normalization was carried out. The mass excision of phagemid library and normalization were done in HT Nguyen lab by D. Zhang at Texas Tech University. Normalization protocol used was that of Soares'. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

ORIGIN

```

Query Match      38.1%; Score 408; DB 13; Length 731;
Best Local Similarity 73.5%; Pred. No. 7e-88;
Matches 535; Conservative 0; Mismatches 190; Indels 3; Gaps 1;

QY 232 GGGATCTTCGTGGAGAAFTACGCTTTATCCGGTCCACTAATCGGAAGGATCTGTCTAT 291
Db 4 GGGATCTTCGTGGAGAGCTCGAGCTGTCGGCTCCACCAACCGAGGGGCTCCCTAC 63
QY 292 ACCTAGGAATCAACCAATTCGCGACTGACCTGGGAGGATTCGGACCAATCGCCTT 351
Db 64 CGCTCGGCATCAACCGTTTCGGGACATGAGCTGGGAGGATTCAGGGGAGCGGCTC 123
QY 352 GTGCGGCGCAGAACTGCTCGCGACTGCGCATGGAACCAACCGGTT---TGTCGATGCG 408
Db 124 GCGCGCGCGAGAACTGCTCGCGAGCGTCCGCGACCAACCGGATCGCGAGCGCGCC 183
QY 409 GTGCTTCCTGTAAAGAGGATGGAGGAGCAAGGATAGTAGAGCCCTGTAAAGAGCAA 468
Db 184 GCTCTCCCGGAGACCAAGACTGGAGGAGGATGGGATCGTAGCCCGCGTGAAGAGCAG 243
QY 469 GGAAGCTGTGATCTTGCTGACTTTCAGTACTACTGAGCACTAGAGGCTGTGATATACA 528
Db 244 GGTCACTGTGTTCTGTGTGACCTTCAGCACCTGTTCTCTTGAGGAGCATATACT 303
QY 529 CAGCTAACTGGAAGAGACATATATCTGAAAGCAACTGTGTGAGCTGTGCTTACGCA 588
Db 304 CAGGCACTGGGAAGCTGTCTCTTCTGAGCAGCAGCTGTGTGATGTGTACTGCA 363
QY 589 TTCATATCTTGGATGCAATGGAGGTTTGCTTCCCAAGCCTTTGAATGTTAAGTAC 648
Db 364 TACAATAATTTCCGATGCAATGGAGGCTTACCCTTCCAGGCTTTTGTAGTACATCAATAC 423
QY 649 AATGAGGATCGACACAGACACTTATCCATCTTGTGTCAATGGTATCTGCAAC 708
Db 424 AATGAGGCTTGACACTGAAAGAGCTTACCCTTACAGCGGTGTCAATGGCATCTGTCAT 483
QY 709 TTCAGCAGGAGATGTTGGTGTCAAGTCAATGATTCGATTAACATCACCTGGTGCT 768
Db 484 TACAAGCCTGAAACGTTGGAGTCAAGGTTTGGACTCCGTTAATCAATCACCTGGTGCT 543
QY 769 GAGGATGAGTTGAAGCATGCAAGTGGCTTGGTGGCTCCAGTTAGCGTTGCAATTTGAGGTT 828
Db 544 GAGGATGAGCTGAAGATGCTGTTGAGCTTGTGTCAGTTAGTGTGCTTTCAGGTG 603
QY 829 GTGAAGGTTTCAATCTGTACAGAAAGTGTATACAGCAGTGAACCTGTGGAGAGAT 888
Db 604 ATCAACGCTTTTCAAGATGAGAGTGGAGTTTACACAAGTGAACCTGTGGAACTTCT 663
QY 889 CCAATGGATGAAACACGAGTCTTTCGCTGCTGATGAGTTCGAGGAGCGGATTCCT 948
Db 664 CCAATGGATGAAACACGCTGTTCTGCGCTTGGCTATGTTGTGAAATGGCGTTCCC 723
QY 949 TATTGGCT 956
Db 724 TACTGGCT 731

```

RESULT 8

CB982399 748 bp mRNA linear EST 01-MAY-2003
LOCUS CAB70006 IvaF A04 Cabernet Sauvignon Berry Post-Veraison - CAB7
DEFINITION Vitis vinifera cDNA clone CAB70006 IvaF_A04 5', mRNA sequence.

ACCESSION CB982399
VERSION CB982399.1 GI:30305805

KEYWORDS

SOURCE Vitis vinifera

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; Vitaceae; Vitis.

REFERENCE

1 (bases 1 to 748)
Goes da Silva,F., Iandolino,A., Lim,H., Baek,J., Jones,K. and
Cook,D.

AUTHORS

Expressed sequence tags from cabernet sauvignon berries at various
developmental stages

JOURNAL

Unpublished (2003)

COMMENT

Contact: Douglas Cook, PhD
CABES Genome Facility
UC Davis, Plant Pathology
One Shields Ave, Davis, CA 95616, USA
Tel: 530 754 6561
Fax: 530 754 6617
Email: drcook@ucdavis.edu

Seq primer: ACGTACCGACATATGCC.

FEATURES

source

```

1..748
/organism="Vitis vinifera"
/mol_type="mRNA"
/cultivar="Cabernet Sauvignon"
/db_xref="taxon:29760"
/clone="CAB70006.IvaF_A04"
/dev_stage="Post-Veraison, 18-19 brx"
/lab_host="DH5alpha"
/clone_lib="Cabernet Sauvignon Berry Post-Veraison - CAB7"
/notes="Organ: Berry; Vector: pDNR; Site 1: SfiI; Site 2: SfiI; CAB7 is a cDNA library of Cabernet Sauvignon clone 8 berries. Samples were collected post-veraison from field-grown vines during stage III of berry growth at 89 days after full bloom. Berries soluble solid concentration ranged between 18-19 brx. Sampled vines were located at the University of California, Davis, Experimental Vineyard. cDNAs were made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows: 5'-AAGCAGGTGATCAACGAGTGGCATTACGCCGG-3' and 5'-ATTCTAGAGCGAGCGCCACATG-3' (30)NN-3'. Library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5-3 kb size fraction."
```

ORIGIN

```

Query Match      38.1%; Score 407.8; DB 14; Length 748;
Best Local Similarity 71.6%; Pred. No. 7.9e-88;
Matches 535; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY 214 GAGATCAAGAAGGTTCCGGATCTTCGTGGAGAACTCTAGCGTTTATCCGGTCCACTAAT 273
Db 2 GAGATTAAGCTGAGATTCGAGATTTTTCGGAGAAATTGAAATCATCATGATCCACCAAC 61
QY 274 CGGAAGATCTGTCTGTATACCTTAGGAATCAACCAATTCGCCGACCTGACCTGGGAGAA 333
Db 62 AGAAAGGCTTGCCTTTATCTCTAGCTGTAAATCAGTTCGCTGATTGGACCTGGGAGAG 121
QY 334 TTCGGACCAATCGCCTTGTGTCGGCGCAGAACTGCTCGCGCACTCGCATGGAACCAAC 393
Db 122 TTCGGCAGACACAGGTTGGAGCTGCTCAGAACTGCTCTGCCACCTTGAAGGCAATCAC 181
QY 394 CGGTTTGTGATGGCTGCTTCTGTGACGAGGATGGAGGAGCAAGGATATGATGAGC 453
Db 182 AAGCTAACTGAAGTTATCTCTGAGACGAAGATTGGAGAGAGATGGCATATATTCAGC 241
QY 454 CTGTGAAGGACCAAGAGCTGTGATCTTCGTGACTTTCAGTACTTCTGAGACTA 513
Db 242 CCAATCAAGATCAAGTCACTGTGATCTTGTGAGCTTTTTCAGCACCCTGAGGCTCTA 301
QY 514 GAGGCTCATATACACAGCTAACTGAAAGAGACATCATATTATCTGAACAGCAACTGTG 573

```



```

Db 302 GAGGACGCTAGCGTCAGGCAATTTGGGAAGGGGATCTCTCTCTGAGCAGCAGCTTGTG 361
QY 574 GACTGTGCTCAGCATCAATCAATTTGATGATCAATGAGGTTTGCCTTCCCAAGCCTTT 633
Db 362 GACTGTGCTCAGCATCAATCAATTTGATGATCAATGAGGTTTGCCTTCCCAAGCCTTT 421
QY 634 GAATAGCTTAAGTACATGAGGATCGACACAGACAGATTAATCCATACCTTGGTGTG 693
Db 422 GAGTACATCAATCAATGATGAGGATCGACACAGACAGATTAATCCATACCTTGGTGTG 481
QY 694 ATGGTATCTGCACTTCAAGCAGGAGATTTGATGATCAATGAGGTTTGCCTTCCCAAGCCTTT 753
Db 482 GATGGCAGCTTCAATCAATTTGATGATCAATGAGGTTTGCCTTCCCAAGCCTTT 541
QY 754 ATCACTCTGGTCTGAGGATGAGTTGAAGCATGACAGTGGGCTTGGTGGCTCCAGTTAGC 813
Db 542 ATTACCTCTGGTCTGAGGATGAGTTGAAGCATGACAGTGGGCTTGGTGGCTCCAGTTAGC 601
QY 814 GTTGCAATTTGAGTTGAAAGTTTCAATCTGTACAGAAAGGTTTATACAGCAGTGAC 873
Db 602 GTGGCAATTTGAGTTGAAAGTTTCAATCTGTACAGAAAGGTTTATACAGCAGTGAC 661
QY 874 ACCTGTGGAAGACATCAATGAGTGTGAACACGACAGTTCTTGGCTCGGTTATGAGTGC 933
Db 662 ACTTGTGGCAGCACTCCAGTGTGACCACTGCTTCTTGCAGTCCGGTATGAGTGT 721
QY 934 GAGGACGGGATCTCTTATTTGGCTCATC 960
Db 722 GAAGATGGTGTAGCATATGGCTCATC 748

```

```

RESULT 9
CK268548 936 bp mRNA linear EST 12-DEC-2003
LOCUS EST714626 potato abiotic stress cDNA library Solanum tuberosum cDNA
DEFINITION clone POACH43 5' end, mRNA sequence.
ACCESSION CK268548
VERSION CK268548.1 GI:39825526
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamiales; Solanales; Solanaceae; Solanum.
1 (bases 1 to 936)
Buell, C.R., Hart, A., Zismann, V., Karamycheva, S.A. and Baker, B.
Generation of ESTs from abiotic stressed potato tissue
Unpublished (2003)
Other ESTs: EST714627
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
Clones can be requested from TIGR via potato@tigr.org
Seq primer: Art TAG GTG ACA CTA TAG.
Location/Qualifiers
1..936
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="POACH43"
/tissue_type="abiotic stress treated leaf and root tissue"
/lab_host="DH10B-Tona"
/clone_lib="potato abiotic stress cDNA library"
/notes="Vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;
supplier: Solanum tuberosum var. Kennebec plants were
grown from cuttings on a 16hr light/8 hr dark cycle at 25
C for 3-4 weeks. Abiotic stress conditions were applied to
four separate sets of plants. Set 1 involved saturation of
the soil with 150 mM NaCl and tissues were harvested at
following application of the salt stress (leaves: 2hr,

```

ORIGIN

```

Query Match 37.9%; Score 405.6; DB 14; Length 936;
Best Local Similarity 68.0%; Pred. No. 3e-87;
Matches 562; Conservative 0; Mismatches 264; Indels 0; Gaps 0;
QY 107 CTGACTGATGAGCCTGCCATCTCGGCGTCTCTTGCAGTTCGCGCAGCCTTCCT 166
Db 102 CCGAAGAGCTGGAGACGGAAATCTTCAAGTGTGCGCCAGACTCGGAATGCTCTCTCT 161
QY 167 TCGCACGCTTCGCGCGAGGTACGGGAAGAGCTACGGATCGGAGGAGGATCAAGAAGA 226
Db 162 TCGTCTGCTTGTCTATCAGGCATCGGAAGAGTACGAGTCCGTTGAGGAGATCAAGCAA 221
QY 227 GGTTCGGGATCTTCGTGGAGAACTAGCGTTTATCCGGTCCACTAATCGGAGGATCTGT 286
Db 222 GGTTCGAGATATTTTGGACAACTGAAGATGATCCGATCGCATAAACAGCAAGGACTAT 281
QY 287 GGTATACCTAGGAATCAACCAATTCGCGACCTGACCTGGGAGGAATTCGCGACCAATC 346
Db 282 CATCAAACTCGGTGTCAATGAGTTTACCGACCTAACATGGATGAGTTCGTAGACACA 341
QY 347 GCCTTGTGTGCGCGCAGAACTGCTCGGGAATCGGCACTGGAAACCCCGGTTTTCGATG 406
Db 342 AGTTGGGGGCACTCAAAACTGTTCTGCCACTACAAAGGGCAATCTCAAGTAACTAACG 401
QY 407 CGCTGCTTCTGTAAACAGGATTTGGGGAGCAAGGATAGTAGGCCCTCTAAAGGACC 466
Db 402 TCGTCTTCTCCAGAGACGAGGACTGGAGGAGATGTTATTTAGCCACTGAGGCAC 461
QY 467 AAGGAAGCTGTGATCTTGTCTGGAATTTAGTACTATCTGGAGCACTAGAGCTGCATATA 526
Db 462 AGGGCAAGTGTGGATCTTGTCTGGACATTCAGCACTACTGTGTCACCTAGAGCAGCATATG 521
QY 527 CACAGCTAACTGGAAAGGACATCATTTATCTGAACAGCACTTGTGAGCTGTGCCTCAG 586
Db 522 CCCAAGCAATTTGGGAAGGGAATCTCTGTGACAGCAGCAGCTTGTGAGCTGTGCTGGAG 581
QY 587 CATTTCAATAAATTTGGATGCAATGAGGTTTGCCTTCCCAAGCCTTTGAAATACGTTAAGT 646
Db 582 CTTTTTAATAAATTTGGCTGCAATGGGGGTTGCCATCACAAAGCCTTTGAGTACATTAAT 641
QY 647 ACAATGGAGGATCGACACAGAAAGATTTATCCATACCTTGTGTCAATGCTATCTCA 706
Db 642 TCAATGGTGGTCTTGACACTGAAGAGCATATCCATACCCGCAAGAAATGGCATATGTA 701
QY 707 ACTTCAAGCAGGAGAAATGTTGTGTCAAGGTCATTTGATTCGATAAACATACCTTGGGTG 766
Db 702 AATTTCTACAGCAATATTTGGTGTCAAGTCATCAGTTCTGTCAATATTTACCTTGGGTG 761
QY 767 CTGAGGATGAGTTGAAGCATCAGTGGGCTTGGTGGCTCCAGTTAGCCTTGCATTTGAGG 826
Db 762 CTGAAGATGAATGAATAACCGCAGTTGCAATTTGGTTAGCCCTGTGTAGTTTGTCTTTGAGG 821
QY 827 TTGTGAAAGTTTCAATCTGTACAGAAAGGTTATACAGCAGTGCACACCTTGTGGAAGAG 886
Db 822 TGTATAAAGGTTTCAACACAGTATAGAGCGGAGTTTACACAGCAGCTGAATGTGGCGACA 881

```

FEATURES

source

/note="Organ: Bud; Vector: pDNR; Site_1: Sfil; Site_2: Sfil; CASBUD is a cDNA library of *Vitis vinifera* cv. 'Cabernet Sauvignon' clone 8 dissected buds. Samples were collected May 13, 2002 from pre-bloom plants (10-11 days before bloom), pre-veraison. Sampled vines were located at the University of California, Davis, Experimental Vineyard. cDNAs were made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:
5' -AAGCATGTGTATCAACGACGAGTGCCATTACGGCGGG-3' and 5' -ATTTTAGCGGAGGCGCCGACATG-dT(30)NN-3'. Library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5-3 kb size fraction."

ORIGIN

```
Query Match          37.6%; Score 402.8; DB 14; Length 848;
Best Local Similarity 76.0%; Pred. No. 1.4e-86;
Matches 497; Conservative 0; Mismatches 157; Indels 0; Gaps 0;
```

412	QY	CTTCTCTGTAAACGAGGGATTTGAGGAGCAAGGATAGTGAGCCCTGTAAAGCAACCAAGGA	471
2	DB	CTTCTCTGAGACGAAGAATTTGAGAGAGAAGATGGCATAGTCAGGCCAAATCAAGAGATCAAGGT	61
472	QY	AGCTGTGGATCTTGTCTGACATTTTCAGTACTACTGGAGCACTAGAGGCTGCATATACAG	531
62	DB	CACCTGTGGATCTTGTCTGACATTTTCAGCACCACTTGGAGCTCTAGAGGCACTTTACGCTCAG	121
532	QY	CTAACTCGGAAGAGACACATCATTTATCTTGAACACGCACTTGTGAGCTGTGCCCTCAGCATTC	591
122	DB	GCATTTGGGAAGGGGATCTCTCTCTGTGAGCAGCAGCTTGTGAGCTGTGCCCGAGCTTTC	181
592	QY	AATAACTTTTGGATGCAATGGAGGTTTGCCTTCCCAAGCCTTTGAAATACGTTTAAGTACAAT	651
182	DB	AATAACTTTTGGGTGCCACCGGGGATGGCCATCCCAAGCTTTTCAGTACATCAAAATACAAT	241
652	QY	GGAGGCATCGACACAGACAGCATTTCCATACCTTGGTGTCAATGGTATCTCGCAACTTC	711
242	DB	GGTGGCCTTTGATCTCTGAGAGAGCATATCCTTACACTGGACTAGTGTGCACTTGCATAAT	301
712	QY	AAGCAGGAGAAATTTTGGTGTCAAGGTCAATGATTCGATAAACATCACCTCGGGTGCTGAG	771
302	DB	TCTTCAGAAAAATTTTGGTGTTCAAAGTTCTCGACTCTGTGAATATTACCTCGGGTGCTGAA	361
772	QY	GATGAGTTCAAGCATGCAGTGGGCTTGGTGGCTCCAGTTAGCGTTTGCATTTTCAGGTTGTG	831
362	DB	GATGAGTTAAAGCATGCAAGTTGCATTTGTTGGTCCAGTGAGTGTGGCAATTTGAGGTGCTC	421
832	QY	AAAGGTTTCAATCTCTGTAACAAGAAAGTGTTATACAGCAGTGCACACTGTGTGAAGAGATCA	891
422	DB	CATGATTTCCGAGTTTTTACAAGAAAGGAGTTTTCACAACAAGTGGAACTTTGTGGCAGCATCC	481
892	QY	ATGGATGTCAACACAGCAGTTCTTGTGCCGTCCGTTATGGAGTCAGAGACGGGATTCCTTAT	951
482	DB	ATGATGTGAACCATGCTGTTCTTGGAGTATGGAGTTGAAGATGGTGTAGCATAC	541
952	QY	TGGCTCATCAAGAACTCATTGGGGTACAAATTTGGGGTGACAAATGGCTCATTTTAAGATGGA	1011
542	DB	TGGCTCATCAAGAACTCATTGGGGAGAAACCTGGGGCGACAAATGGCTCATTTCAAAACGGAG	601
1012	QY	CTCGGCAAGAAATGTTGGTGTTCGCAACTTTCGCGCATCTTATCCCATTTGGCT	1065
602	DB	TTGGGCAAGAAATGTTGGTGTTCGCAACTTGTCTCATCATATCCGTTTGTGCT	655

REC'D JUL 12

RESULT 12
CF510584/C

STOCKS

CF510584 886 bp mRNA linear EST 09-SEP-2003

Accession	Parent	Sex	Year	Location	Notes
CBud00001	TR F08	Vitis vinifera cv. cabernet sauvignon	1988	Clon 8	

Bud - CABUD Vitis vinifera cDNA clone CABUD0001 IR F08 3', mRNA

sequence
and - LAB

ACCESSION

ACCESS
WEBSITE

CF510584 1 CT.34542352

55

SOURCE ORGANISM	Vitis vinifera Vitis vinifera Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; rosids; Vitaceae; Vitis.
REFERENCE	1 (bases 1 to 888)
AUTHORS	Goes da Silva,F., Iandolo,A., Lim,H., Baek,J., Jones,K. and Cook,D.
TITLE	Expressed sequence tags from Vitis vinifera 'Cabernet sauvignon' berries at various developmental stages
JOURNAL	Unpublished (2003)
COMMENT	Contact: Douglas Cook, PhD CAES Genome Facility UC Davis, Plant Pathology One Shields Ave, Davis, CA 95616, USA Tel: 530 754 6561 Fax: 530 754 6617 Email: drcook@ucdavis.edu Seq primer: GCCAACGGAATGGCTAG.

FEATURES
SOURCE

```

1. .886
/organism="Vitis vinifera"
/mol_type="mRNA"
/cultivar="Cabernet Sauvignon (Clone 8)"
/db_xref="taxon:29760"
/clone="CABud0001_IG_F08"
/sex="Hermaphrodite"
/dev_stage="Pre-bloom (10-11 days before bloom)"
/lab_host="DH5alpha"
/clone_lib="Vitis vinifera cv. cabernet sauvignon (Clone
8) Bud - CABUD"
/notes="Organ: Bud; Vector: pDNR; Site1: Sfil; Site2:
Sfil; CABUD is a cDNA library of Vitis vinifera cv.
'Cabernet Sauvignon'. Clone 8 dissected buds. Samples were
collected May 13, 2002 from pre-bloom plants (10-11 days
before bloom), pre-veraison. Sampled vines were located at
the University of California, Davis, Experimental
vineyard. cDNAs were made by oligo-dT priming and
directionally cloned. 5' and 3' adaptors were used in
cloning as follows:
5'-AAGCATGGTGATCAACGCGGCAGTCATTACGCCGGG-3' and
5'-ATTCTAGGCGGCGAGCGGCAGCAT-dt(30)NN-3'. Library was
constructed using the Clontech Creator SMART kit and
size-selected to contain the 0.5-3 kb size fraction."

```

ORIGIN

Query Match	37.5%	Score	401.8	DB	14	Length	886
Best Local Similarity	76.0%	Pred.	No. 2.4e-86				
Matches	496	Conservative	0	Mismatches	157	Indels	0
QY	413	TTCTCTGAACGAGGGATTGGAGGGAGCAAGGGATAGTGGCCCTGTAAAGAGCAACCAAGAA	472				
DB	886	TTCTCTGAGACGAAGATTTGGAGAGGAATGGCATAGTCAGCCCAATCAAAGATCAAGGTC	827				
QY	473	GCTCTGATCTTGTCTGACCTTTCAGTACTACTTGGAGCACTAGAGGCTGCGATATACACAGC	532				
DB	826	ACTGTGATCTTGTCTGACCTTCAGCACCACTGGAGCTCTAGAGCGAGCTTACGCTCAGG	767				
QY	533	TAACTGAAAGAGACATCATATCTGAAACAGCAACTGTGGACTGTGCTCAGCATTCAC	592				
DB	766	CATTGGGAAGGGATCTCTGCTCTGAGCAGCAGCTTGTGGACTGTGCCGAGCTTTCA	707				
QY	593	ATACTTTGGATGCAATGGAGGTTTGCTTCCCAAGCCTTTTGAATACGTTAAAGTACAATG	652				
DB	706	ATACTTTGGGTGCCACGGGGATGGCATCCCAAGCTTTTGAATACATCAATACAAATG	647				
QY	653	GAGGCATCGACACAAACAGACTTATCCATACCTTTGTTGTCAAATGGTATCTGCAACTTCA	712				
DB	646	GTGGCCTTGATCTAGGAAGCATATCCTTACACTGGACTAGATGGCACCTGCAAAATTTT	587				
QY	713	AGCAGGAGAAATGTTGGTGTCAAGGTCAATTCGATTAACATCACCCCTGGGTGCTGAGG	772				
DB	586	CTTCAGAAATATTGGTGTTCAGCTCTCGACTCTGTAATATTACCTGCTGCTGCTGAAG	527				

RESULT 12	CF510584	886 bp	linear	EST 09-SEP-2003
LOCUS	CABud0001_IR_F08			
DEFINITION	Bud - CABUD <i>Vitis vinifera</i> cv. cabernet sauvignon (Clone 8)			
	Bud - CABUD <i>Vitis vinifera</i> cdna clone CABud0001_IR_F08 3', mRNA			
	sequence.			
ACCESSION	CF510584			
VERSION	CF510584.1			
KEYWORDS	GI:34542352			
	EST.			

ORIGIN	Query Match	Score	DB	Length
	37.2%	398.6	14	977

RESULT 14	720 bp	mRNA	linear	EST 20-FEB-2001
BE131652				
LOCUS				
DEFINITION				
	BE131652			
	L48-1651/3	ice plant	Lambda Uni-Zap XR expression	library, 48 hours
	NaCl treatment	Mesembryanthemum	crystallinum	cDNA clone L48-1651,

```

mRNA sequence.
ACCESSION BE131652
VERSION BE131652.1 GI:8579015
KEYWORDS EST.
SOURCE Mesembryanthemum crystallinum (common iceplant)
ORGANISM Mesembryanthemum crystallinum
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
TITLES Caryophyllales; Alzooaceae; Mesembryanthemum.
JOURNAL 1 (bases 1 to 720)
COMMENT Cushman, J.C.
An expressed sequence tag database for the common ice plant,
Mesembryanthemum crystallinum
Unpublished (1997)
Contact: Cushman JC
Department of Biochemistry
University of Nevada
MS200, Reno, NV 89557-0014, USA
Tel: 775-784-1918
Fax: 775-784-1650
Email: jcushman@unr.edu
PCR Primers
FORWARD: T7
BACKWARD: T3
Plate: L48-17 row: E column: 3
Seq primer: T3
High quality sequence stop: 350
POLYA=No.

FEATURES             Location/Qualifiers
     source            1..720
     mol_type="mRNA"   /organism="Mesembryanthemum crystallinum"
     db_xref="taxon:3544"
     /clone="L48-1651"
     /tissue_type="Leaf, 48 h 0.4M NaCl"
     /dev_stage="Six week old"
     /clone_lib="Ice plant Lambda Uni-Zap XR expression
     library, 48 hours NaCl treatment"
     /note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
     EcoRI; Site_2: XhoI"

ORIGIN
Query Match      36.8%; Score 394.2; DB 10; Length 720;
Best Local Similarity 72.2%; Pred. No. 1.5e-84;
Matches 513; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

QY 350 TTGGTGGCGGCAGAACTGCTCGCGAGTGGCGATGGAACCCGCTTTGTTCGATGGCG 409
DB 10 TGGGAGCTGCCAAAACCTGCTCTGCCACTCTCAAGGCGAGTCACAGCTCTCTCATGCGAG 69
QY 410 TGCTTCTGTAAACGAGGATTTGAGGAGCAAGGATAGTGCAGCCCTGTAAAGGACCAAG 469
DB 70 ACCTTCTGAAACAAAGATTTGAGAGCTGAGGCGCATAGTCAGCCCGGTCAAGATCAAC 129
QY 470 GAAGCTGTGGATCTTGTGTGACTTTCAGTACTTGGAGCACTAGAGGCTGCATATACAC 529
DB 130 ACCACTGTGGATCTCTGTGGACATTCAGTACAACTGGAGCTTTGGAGCGAGCTTATGCGC 189
QY 530 AGCTAACGGAAGAGCAGACATCATTTATCTGAACAGCAACTGTGCACTGTGCTTCAGCAT 589
DB 190 AGGCATTTGGAGAGAACATCTTTTGTCTGAGCAACACACTTGTGATTTGCTGTGTCAT 249
QY 590 TCAATAACTTTTGGATGCAATGGAGTTTGCCTTCCAGCCCTTTGAAATAGTTAAGTACA 649
DB 250 TTGATAATCATGGTTGCAATGGTGGTTGGCATCCCAAGCTTTTCAGTATGTCAAGTACA 309
QY 650 ATGGAGGATCGACACAGCAAGACTTATCATCTTGGTGTGTCATGGTATCTGCAACT 709
DB 310 ATGGAGGACTTGACACAGAGAGGCTTATCTTACACAGCAAGAGTGGTGAATGTAAT 369
QY 710 TCAAGCAGGAGAAATGTTGGTGTCAAGGTCAATGATTCGATAAAACATCAACCTGGGTGCTG 769
DB 370 TCTCAGCCCAAAATGTTGGTGTTCAGTCTCGTGGACTCGGTAAACATCACCCCTCGGAGCTG 429

```

```

770 AGGATGAGTTGAAGCATGCACTGGGCTTGGTGGTCCAGTACGCTTGCATTTGAGGTTG 829
DB 430 AAGAGGAACCTTAAGGATGCTGTGGCAATGGTCCGCTCTGTGAGTGTGCGATTTGAGGTTG 489
QY 830 TGAAGGTTTCAATCTGTACAAGAAAGGTGTATACAGCAGTGACACCTGTGGAAGAGATC 889
DB 490 TTATGATTTCCGGTTTACACGAAGAGGATTTACACCAGCAACACTTTGTGGCAGCAGCC 549
QY 890 CAATGGATGTGAACACGCGAGTTCTTCCCGCTGTTATGGAGTCGAGGAGGATTCCTT 949
DB 550 CCATGGATGTGAACCATGCTGTTTCTAGCTGTGGATATGGATGGAGATGGTATCCCAT 609
QY 950 ATTGGCTCATCAAGAACTCATGGGTACAAATTTGGGTGACAATGGCTACTTTAAGATGG 1009
DB 610 ACTGGCTTATCAAGAACTCTGTGGGAGAGGATTTGGGTGACAATGGATCTTCAAGATGG 669
QY 1010 AACTCGGCAAGAACATGTGTGGTTTGCACCTTGGCATCTTATCCCATGG 1060
DB 670 AGATGGGAAAAAACATGTGTGGTGTCTACTTGTGGCGCATATCCAGTTG 720

RESULT 15
CK294622      965 bp      mRNA      linear      EST 15-DEC-2003
LOCUS      EST757336 Nicotiana benthamiana mixed tissue cDNA library,
DEFINITION      normalized, full-length Nicotiana benthamiana cDNA clone NEMC076 5'
end, mRNA sequence.
ACCESSION      CK294622
VERSION      CK294622.1 GI:39878196
KEYWORDS      EST.
SOURCE      Nicotiana benthamiana
ORGANISM      Nicotiana benthamiana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamiales; Solanales; Solanaceae; Nicotiana.
1 (bases 1 to 965)
Buell, C.R., Hart, A., Zismann, V., Karamycheva, S.A., Day, B.,
Staskawicz, B., Jin, H. and Baker, B.
Generation of EST sequences from Nicotiana benthamiana
Unpublished (2003)
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
Clones can be requested from TIGR via potato@tigr.org
Seq primer: ATT TAG GTG ACA CTA TAG.

FEATURES             Location/Qualifiers
     source            1..965
     organism="Nicotiana benthamiana"
     mol_type="mRNA"
     db_xref="taxon:4100"
     /clone="NEMC076"
     /tissue_type="abiotic and biotic stress-treated leaves,
     callus tissue and root tissue"
     /lab_host="DH10B-TonA"
     /clone_lib="Nicotiana benthamiana mixed tissue cDNA
     library, normalized, full-length"
     /note="Vector: pCMVSPORT6.1; Site_1: EcoRI; Site_2: NotI;
     supplier: RNA was isolated from Nicotiana benthamiana
     tissues that include callus, roots from liquid culture
     grown plants, heat-stressed leaves (38 C, 3 hr and 6 hr),
     cold-stressed leaves (5 C 3 hr, 6hr), and pathogen
     challenged leaves (Pseudomonas syringae pv tomato 12 hr;
     Xanthomonas campestris pv campestris 12 hr, 18hr;
     Pseudomonas syringae pv phaseolicola 18hr, and Xanthomonas
     campestris pv vesicatoria 18hr). RNA was isolated from
     these tissues and pooled in approximately equal molar
     amounts."

ORIGIN
Query Match      36.5%; Score 390.8; DB 14; Length 965;
Best Local Similarity 64.9%; Pred. No. 1.2e-83;

```


PT supplementing tissue (e.g., embryo) culture of *V. planifolia* with malic
 PT acid or by subjecting the culture to heat or mechanical shear stress.
 XX Claim 18; Fig 6; 49pp; English.
 XX The present invention relates to a method for improving vanillin
 CC production in cultured *Vanillin planifolia*, and in intact plants. The
 CC method involves genetically engineering *V. planifolia* to overproduce
 CC enzymes associated with step(s) involved in vanillin biosynthesis in the
 CC plant. Transgenic cells and plants overproducing 4-hydroxybenzaldehyde
 CC synthase (4HBS) enzyme are also provided. The method is useful for
 CC improving vanillin production in cultured *V. planifolia*. The method
 CC results in the production of cultured cells or plants producing at least
 CC twice to 10 times more vanillin than cells or plants not produced by the
 CC new method. The present sequence represents *V. planifolia* 4HBS.
 XX Sequence 352 AA;

Query Match 100.0%; Score 1865; DB 7; Length 352;
 Best Local Similarity 100.0%; Pred. No. 5.6e-176; Indels 0; Gaps 0;
 Matches 352; Conservative 0; Mismatches 0;
 QY 1 MAAKLFFLLVLSVALAGFEEDNPIRSVTORPDSIEPAIILGVLSGRHAFHAFPA 60
 DB 1 MAAKLFFLLVLSVALAGFEEDNPIRSVTORPDSIEPAIILGVLSGRHAFHAFPA 60
 QY 61 RYKSGYSGSEIEKKRFGIFVENLAFIRSTNRKDLSTLGINQADLTWEEFTNRLGAA 120
 DB 61 RYKSGYSGSEIEKKRFGIFVENLAFIRSTNRKDLSTLGINQADLTWEEFTNRLGAA 120
 QY 121 QNCSTAHGNHRFVDGVLPTVRDREOGIVSPVKDQSGSWTSTTGAAEAAYTQTGS 180
 DB 121 QNCSTAHGNHRFVDGVLPTVRDREOGIVSPVKDQSGSWTSTTGAAEAAYTQTGS 180
 QY 181 TLSEQQLVDCASAFNFCGGLPSQAFYVYKNGIDTEQTPYLVGMGICNFKQENVG 240
 DB 181 TLSEQQLVDCASAFNFCGGLPSQAFYVYKNGIDTEQTPYLVGMGICNFKQENVG 240
 QY 241 KVIDSINITLGADELKHAAGLVPRVSVAFVWKGFLYKKGVSSTTCGRDPMVNHAV 300
 DB 241 KVIDSINITLGADELKHAAGLVPRVSVAFVWKGFLYKKGVSSTTCGRDPMVNHAV 300
 QY 301 LAVGYGVEDGIPYWLKNSWGTNGDNGYFKWELGKNCMGVATCASYP 352
 DB 301 LAVGYGVEDGIPYWLKNSWGTNGDNGYFKWELGKNCMGVATCASYP 352

RESULT 2
 ADC61399
 ID ADC61399 standard; protein; 363 AA.
 XX
 AC ADC61399;
 XX
 DT 18-DEC-2003 (first entry).
 XX
 DE Zea mays cysteine protease #2.
 XX
 KW Vanillin production; plant; vanillin biosynthesis;
 KW 4-hydroxybenzaldehyde synthase; 4HBS; enzyme; cysteine protease.
 XX
 OS Zea mays.
 XX
 PN US2003070188-A1.
 XX
 PD 10-APR-2003.
 XX
 PF 28-FEB-2002; 2002US-00087714.
 XX
 PR 15-JUL-1997; 97US-0052604P.
 PR 15-JUL-1998; 98WO-US014895.
 PR 25-MAY-2000; 2000US-00462576.
 PR 28-FEB-2001; 2001US-0272415P.
 XX

PA (HAYK/) HAVKIN-FRENKEL D.
 PA (PODS/) PODSTOLSKI A.
 PA (DIXO/) DIXON R A.
 XX
 PI Havkin-Frenkel D, Podstolski A, Dixon RA;
 XX WPI; 2003-743794/70.
 XX
 DR Improving vanillin production in cultured *Vanillin planifolia* by
 XX supplementing tissue (e.g., embryo) culture of *V. planifolia* with malic
 PT acid or by subjecting the culture to heat or mechanical shear stress.
 PT Disclosure; Fig 6; 49pp; English.
 XX
 PS The present invention relates to a method for improving vanillin
 CC production in cultured *Vanillin planifolia*, and in intact plants. The
 CC method involves genetically engineering *V. planifolia* to overproduce
 CC enzymes associated with step(s) involved in vanillin biosynthesis in the
 CC plant. Transgenic cells and plants overproducing 4-hydroxybenzaldehyde
 CC synthase (4HBS) enzyme are also provided. The method is useful for
 CC improving vanillin production in cultured *V. planifolia*. The method
 CC results in the production of cultured cells or plants producing at least
 CC twice to 10 times more vanillin than cells or plants not produced by the
 CC new method. The present sequence represents a plant cysteine protease
 CC used for comparison to the *V. planifolia* 4HBS protein of the invention.
 XX Sequence 363 AA;

Query Match 70.9%; Score 1323; DB 7; Length 363;
 Best Local Similarity 70.8%; Pred. No. 3.3e-122;
 Matches 257; Conservative 30; Mismatches 64; Indels 12; Gaps 6;
 QY 1 MAAKLFPFLPL-----VSALSVALAGFEEDNPIRSVTORPDS-IPAILGVLSGRHAF 54
 DB 1 MAHRRIILLAVAATAVSAVAASGCFDSDNPIRPTDRAASALESTVPAALGRTRDAL 60
 QY 55 HPAPFARYKSGYSGSEIEKKRFGIFVENLAFIRSTNRKDLSTLGINQADLTWEEFT 114
 DB 61 RFAPFAVRYKSGYSGSEIEKKRFGIFSESLQVLRSTNRKGLSYRLGVNRFADMSWEEFRA 120
 QY 115 NRLGAQNCSTAHGNHRF--VDGVLPTVRDREOGIVSPVKDQSGCS-WTSTTGAE 171
 DB 121 TLGNAQNCSTALTGNERMRAAVALPETKDWREDGIVSPVKQHGCSWTSTTGAE 180
 QY 172 AAYTQLTGS--TLSEQQLVDCASAFNFCG-GLPSQAFYVYKNGIDTEQTPYLVGM 228
 DB 181 AAYTQATGKPTSLSEQQLVDCGFAFNFCGNGGLPSQAFYIYKNGGLDTEESYPYQGVN 240
 QY 229 GICNFKQENVGVKVIDSINITLGADELKHAAGLVPRVSVAFVWKGFLYKKGVSST 288
 DB 241 GICNFKQENVGVKVIDSINITLGADELKHAAGLVPRVSVAFVWKGFLYKKGVSST 300
 QY 289 CGRDPMDVNHAVLAVGYGVEDGIPYWLKNSWGTNGDNGYFKWELGKNCMGVATCASYP 348
 DB 301 CGTTPMDVNHAVLAVGYGVEDGIPYWLKNSWGTNGDNGYFKWELGKNCMGVATCASYP 360
 QY 349 IVA 351
 DB 361 IVA 363

RESULT 3
 ADC61398
 ID ADC61398 standard; protein; 360 AA.
 XX
 AC ADC61398;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Zea mays cysteine protease #1.
 XX
 KW Vanillin production; plant; vanillin biosynthesis;
 KW 4-hydroxybenzaldehyde synthase; 4HBS; enzyme; cysteine protease.

XX Zea mays.
 XX OS
 XX US2003070188-A1.
 XX PN
 XX 10-APR-2003.
 XX PD
 XX 28-FEB-2002; 2002US-00087714.
 XX PF
 XX 15-JUL-1997; 97US-0052604P.
 XX PR
 XX 15-JUL-1998; 98WO-US014895.
 XX PR
 XX 25-MAY-2000; 2000US-00462576.
 XX PR
 XX 28-FEB-2001; 2001US-0272415P.
 XX XX
 XX (HVK/) HAVKIN-FRENKEL D.
 XX PA (PODS/) PODSTOLSKI A.
 XX PA (DIXO/) DIXON R A.
 XX XX
 XX Havkin-Frenkel D, Podstolski A, Dixon RA;
 XX PI
 XX WPI; 2003-743794/70.
 XX DR
 XX
 XX Improving vanillin production in cultured Vanillin planifolia by
 PT supplementing tissue (e.g., embryo) culture of V. planifolia with malic
 PT acid or by subjecting the culture to heat or mechanical shear stress.
 PT
 XX
 XX Disclosure; Fig 6; 49pp; English.
 XX PS
 XX
 XX The present invention relates to a method for improving vanillin
 CC production in cultured Vanillin planifolia, and in intact plants. The
 CC method involves genetically engineering V. planifolia to overproduce
 CC enzymes associated with step(s) involved in vanillin biosynthesis in the
 CC plant. Transgenic cells and plants overproducing 4-hydroxybenzaldehyde
 CC plant. (HBS) enzyme are also provided. The method is useful for
 CC synthase (HBS) enzyme are also provided. The method is useful for
 CC improving vanillin production in cultured V. planifolia. The method
 CC results in the production of cultured cells or plants producing at least
 CC twice to 10 times more vanillin than cells or plants not produced by the
 CC new method. The present invention represents a plant cysteine protease
 CC used for comparison to the V. planifolia 4HS protein of the invention.
 XX CC
 XX SQ Sequence 360 AA;
 Query Match 69.8%; Score 1302.5; DB 7; Length 360;
 Best Local Similarity 71.1%; Pred. No. 3.5e-120;
 Matches 256; Conservative 26; Mismatches 69; Indels 9; Gaps 6;
 QY 1 MAKLFLFLVLSALVALA--GPEEDNPIRSVTQRPDS--TEPAILGVLSGRHAFHA 57
 DB 1 MAPRRLLVLAVALAATAAANS GFADSNPIRPVTDRAASALESTVFAALGTRDALRFA 60
 QY 58 RFARRYKSGYSEEEIKKRFIFVENLAFIRSTNRKDLSTYTLGINQFADLTWEEFTRNL 117
 DB 61 RFAVRKGSYESAAEVHKEFRIFSESLQLVRSNKRGLSYRLGINRFADMSWEEFTRNL 120
 QY 118 GAAQNCASATAGNHRF--VDGVLVPTVRDWRBQGIIVSPVKDQSCGS--WTFSTTGALBAAY 174
 DB 121 GAAQNCASATLGNHRRAAVALPETKDRBQGIIVSPVKDQSCGS--WTFSTTGALBAAY 180
 QY 175 TQLTGS--TLSEQLVDCASAFNFC--GGLPSQAFYKYNKGIDTTEQTYPLGVMGIC 231
 DB 181 TQATGPISLSEQLVDCGLAFNFCNGGLPSQAFYKYNKGIDTTEQTYPLGVMGIC 240
 QY 232 NFKQENVGVKIDSINITLGADELKHAVGLVRPVSVAPEVVKGNLYKKGVYSDDTCGR 291
 DB 241 KFNENNVGVKIDSVNITLGADELKDAVGLVRPVSVAPEVITGFRLYKSGVVTSDHCGT 300
 QY 292 DPMNVNHAVALVGYGVEDGIPYWLKINSWGTWNGDGYFKMELGKNMCGVATCASYPIVA 351
 DB 301 TPMNVNHAVALVGYGVEDGIPYWLKINSWGTWNGDGYFKMELGKNMCGVATCASYPIVA 360
 RESULT 4
 AAB65775

AAB65775 standard; protein; 378 AA.
 AAB65775;
 27-MAR-2001 (first entry)
 Cysteine protease #16.
 Cell death modulator; programmed cell death; PCD; apoptosis;
 forestry plant.
 Eucalyptus grandis.
 WO2000075331-A1.
 14-DEC-2000.
 02-JUN-2000; 2000WO-NZ000086.
 04-JUN-1999; 99US-00325932.
 (GENE-) GENESIS RES & DEV CORP LTD.
 (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 Flinn B, Lasham A;
 WPI; 2001-061724/07.
 N-FSDB; AAF44802.
 Novel defender against cell death polynucleotide useful for modulating
 programmed cell death pathway and specific development pathways in
 forestry plant.
 Claim 22; Page 112-113; 142pp; English.
 The present invention relates to coding sequences (see AAF44740-F44840
 and AAF44843-F44844) and proteins (see AAB657714-B65814) involved in
 programmed cell death (PCD; apoptosis). The coding sequences and proteins
 of the present invention are useful for modulating a PCD or cell death
 pathway and various developmental pathways in a forestry plant, by stably
 incorporating one of the present coding sequences into the genome of the
 forestry plant, where the coding sequence provides a PCD pathway that is
 not present in a native form of the forestry plant
 SQ Sequence 378 AA;
 Query Match 69.7%; Score 1300.5; DB 4; Length 378;
 Best Local Similarity 70.4%; Pred. No. 6e-120;
 Matches 247; Conservative 33; Mismatches 60; Indels 11; Gaps 5;
 QY 9 LFLVLSALVALAGFEEDNPIRSVTQRPD----SIEPAILGVLSGRHAFHAFARFARYG 64
 DB 31 LLLVAVVVSAAASFEESNPIRLP---PDGGLRDLESSIVQIVGTRHAPSFARFANRYG 87
 QY 65 KSYGSEEEIKKRGIFVENLAFIRSTNRKDLSTYTLGINQFADLTWEEFTRNLGAAQNC 124
 DB 88 KRYETABEIKLRFEIFRENKLIRSTNKGFLPTLGVNKFADWSWEEFRHRLGAAQNC 147
 QY 125 ATAGNHRFVDGVLVPTVRDWRBQGIIVSPVKDQSCGS--WTFSTTGALBAAYTQLTGS--T 181
 DB 148 ATTGNHKLTDALPEMKOWREKGIIVSPIKQGHGSCSWTFSTTGALBAAYHQAFGKQIS 207
 QY 182 LSEQLVDCASAFNFC--GGLPSQAFYKYNKGIDTTEQTYPLGVMGICNFKQENVGV 240
 DB 208 LSEQLVDCAGAFNFC--GGLPSQAFYKYNKGIDTTEQTYPLGVMGICNFKQENVGV 267
 QY 241 KVIDSINITLGADELKHAVGLVRPVSVAPEVVKGNLYKKGVYSDDTCGRDPMNVNHA 300
 DB 268 QVLDSVNITLGADELKHAVAFVRPVSVAPEVVKGNLYKKGVYSDDTCGSTSDVNHAV 327
 QY 301 LAVGYGVEDGIPYWLKINSWGTWNGDGYFKMELGKNMCGVATCASYPIVA 351
 DB 328 LAVGYGVEDGIPYWLKINSWGTWNGDGYFKMELGKNMCGVATCASYPIVA 378

RESULT 5
ADB94794
XX ADB94794 standard; protein; 378 AA.
XX AC ADB94794;
XX DT 04-DEC-2003 (first entry)
XX DE Programmed cell death pathway cysteine protease #16.
XX KW programmed cell death; plant development; plant cell cycle; ATL2; DAD1;
KW Dnase; lls; lsd; nucellin-like aspartic protease; annexin; prohibitin;
KW fen-like protein; rac2; retinoblastoma-related protein; SINA; rfid;
KW TEGT; xylogenic RNase; pur-alpha; cyteine protease; RPP5-like protein;
KW gp 91 NADPH oxidase subunit; NPR-like protein; BAG-1;
KW defender against cell death; lethal leaf spot; lesion stimulating death;
KW seven in absentia; transcription initiation factor;
KW testis enhanced gene transcript.
XX OS Eucalyptus grandis.
XX PN US2003082724-A1.
XX PD 01-MAY-2003.
XX PF 14-AUG-2002; 2002US-00219220.
XX PR 04-JUN-1999; 99US-00325932.
XX PA (GENE-) GENESIS RES & DEV CORP LTD.
XX PI Flinn B, Lasham A;
XX DR WPI; 2003-786916/74.
XX PT New isolated polynucleotide useful for modulating programmed cell death,
PT altering the development cycle of plant cells, and subsequently modifying
PT plant development.
XX PS Claim 12; Fig 12; 214pp; English.
XX CC The invention describes an isolated polynucleotide (I) comprising a
CC nucleotide sequence that is one of 145 fully defined sequences of 221-
CC 3415 base pairs (bp), given in the specification, and/or its complements,
CC reverse complements, reverse sequences, or sequences having 75, 90 or 95
CC % sequence identity to, or that hybridise under stringent hybridisation
CC conditions to one of the 145 sequences. The methods and compositions of
CC the present invention to do with (I) are useful for modulating programmed
CC cell death and thereby altering the development cycle of plant cells, and
CC altering plant development. This is the amino acid sequence of a protein
CC associated with the programmed cell death pathway.
XX SQ Sequence 378 AA;
Query Match 69.7%; Score 1300.5; DB 7; Length 378;
Best Local Similarity 70.4%; Pred. No. 6e-120;
Matches 247; Conservative 33; Mismatches 60; Indels 11; Gaps 5;
QY 9 LFLVLSALVALAGFEENPFRSVTQRPD---STEPAILGVLSGRHAFHAFARRYG 64
DB 31 LLLVAVVSAASSTFEEENPILF--PDGGLDLESSIVQIVGTRHAFSFAFARRYG 87
QY 65 KSYGSEETIKKFGIFVENLAFIRSTNRKDISYTLGINQFADLTWEEFTRNLGAQNCS 124
DB 88 KRYETAEEIKLRFEIFRENKLIRSTNKKLPYTLGNKFNADWSWEEFRHRLGAQNCS 147
QY 125 ATAHGNHRVDGVLPTWRDRQGIQVSPVKGSCGS-WTSTTGCALEAAVYTLTGS--T 181
DB 148 ATTGNHKLTDALPEMKDWRKGIQVSPVKGSCGSWTFSTTGCALEAAVYTLTGS 207
QY 182 LSEQQLVDCASAFNFGC--GGLPSQAFEVYKNGGIDTEQTYPYLVGMGICNFKQENVGV 240

DB 208 LSEQQLVDCASAFNFGCGLPSQAFEVYKNGGLDTEETAYPTAVDGSCKFSADNVGV 267
QY 241 KVIDSINITLGAEDLKHAVGLVRPVSVAFEVVKFNLYKKGVSSTCGRDPMDVNHAV 300
DB 268 QVLDVSNITLGAEDLKHAVGLVRPVSVAFOVKDFRLYKSGVYTSDTGSGTSMNVNHAV 327
QY 301 LAVGYGVEDGIPYWLKNSWGTNWDNGYFKWELGKMGCGVATCASYPIVA 351
DB 328 LAVGYGVEDGVFFWLLIKNSWGDWDGHDGFKWEMGKMGCGVATCASYPVVA 378
RESULT 6
ADC61397
ID ADC61397 standard; protein; 360 AA.
XX AC ADC61397;
XX DT 18-DEC-2003 (first entry)
XX DE Nicotiana tabacum cysteine protease.
XX KW Vanillin production; plant; vanillin biosynthesis;
KW 4-hydroxybenzaldehyde synthase; 4HBS; enzyme; cysteine protease.
XX OS Nicotiana tabacum.
XX PN US2003070188-A1.
XX PD 10-APR-2003.
XX PF 28-FEB-2002; 2002US-00087714.
XX PR 15-JUL-1997; 97US-0052604P.
PR 15-JUL-1998; 98WO-US014895.
PR 25-MAY-2000; 2000US-00462576.
PR 28-FEB-2001; 2001US-0272415P.
XX PA (HAVK/) HAVKIN-FRENKEL D.
PA (PODS/) PODSTOLSKI A.
PA (DIXO/) DIXON R A.
XX PI Havkin-Frenkel D, Podstolski A, Dixon RA;
XX DR WPI; 2003-743794/70.
XX PT Improving vanillin production in cultured Vanillin planifolia by
PT supplementing tissue (e.g., embryo) culture of V. planifolia with malic
PT acid or by subjecting the culture to heat or mechanical shear stress.
XX PS Disclosure; Fig 6; 49pp; English.
XX CC The present invention relates to a method for improving vanillin
CC production in cultured Vanillin planifolia, and in intact plants. The
CC method involves genetically engineering V. planifolia to overproduce
CC enzymes associated with step(s) involved in vanillin biosynthesis in the
CC plant. Transgenic cells and plants overproducing 4-hydroxybenzaldehyde
CC synthase (4HBS) enzyme are also provided. The method is useful for
CC improving vanillin production in cultured V. planifolia. The method
CC results in the production of cultured cells or plants producing at least
CC twice to 10 times more vanillin than cells or plants not produced by the
CC new method. The present sequence represents a plant cysteine protease
CC used for comparison to the V. planifolia 4HBS protein of the invention.
XX SQ Sequence 360 AA;
Query Match 69.0%; Score 1286; DB 7; Length 360;
Best Local Similarity 69.6%; Pred. No. 1.5e-118;
Matches 247; Conservative 35; Mismatches 65; Indels 8; Gaps 5;
QY 5 LFLFLVLSVALAG---FEEDNPIRS-VTQRPDSIEPAILGVLSGRHAFHAFARFA 60
DB 6 LLLALVVAGLFSALAGPATFADENPIRQVSDGLHELENAILQVVGTRHALSPARFA 65


```

PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150386P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 16-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0156559P.
PR 28-SEP-1999; 99US-0156559P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158023P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159328P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161408P.
PR 25-OCT-1999; 99US-0161406P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.

PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 68.6%; Score 1279.5; DB 3; Length 358;
Best Local Similarity 67.0%; Pred. No. 6.7e-118; Indels 7; Gaps 4;
Matches 240; Conservative 39; Mismatches 72;

QY 1 MAAKLLF---FLPLVLSALVAGFEEDNPFRSVTQRPDSIEPAILGVLSGRHAFHFA 57
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MSAKTLSSVVLVWLVAASAAANIGFDSNPFRVWSDGLREVRSVSQILQSRHVL SFA 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 58 RFARRYKSYGSEBEIKRFGIFVENLAFIRSTNEKDLSTLGNQFADLTWERPTNRL 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 RFTHRYGKYQNVEMKLRFSIFKENLDLIRSTNKKSLGVLGNVQFADLTWQEFQTKL 120
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 118 GAAQNCSATAGNHRFVDGLPVTRDREOGIVSPVKDQSGCS-WTFSTTTGALEAAV 176
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 GAAQNCSATLKGSHKVTETALPETKDWREDGLVSPVKDQSGCSWTFSTTTGALEAAV 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 177 LTGS---TLSEQQLVDCASAFNNFGC-GGLPQAFYVVKXNGGIDTEQTPYLVGMGICNF 233
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 AFGKGISLSEQQLVDCAGAFNNYGCNGLPQAFYIKSNGGLDTEKAYPTGKDETCF 240
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 234 KOENVGVKVIDSINITLGADELKHAVLVRPVSVAFEVWVGFLNLYKKGVYSSDTCGRDP 293
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 241 SAENVGVQVLSNVNITLGADELKHAVLVRPVSIAFEVIHSFRLYKSGVYTDHSGSTP 300
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 294 MDVNHAVLAAGYGVBDGTPYMLIKNSWGTNMGDNGYFKMELKKNMCGVATCASYIVA 351
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 301 MDVNHAVLAAGYGVBDGTPYMLIKNSWGDWGDGKGYFKMGMKKNMCGIATCASYIPVVA 358
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
AAG47118
ID AAG47118 standard; protein; 358 AA.
XX
AC AAG47118;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 59353.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; Genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 05-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 08-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
```

PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137502P.
PR 04-JUN-1999; 99US-0137724P.
PR 07-JUN-1999; 99US-0138094P.
PR 08-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140693P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0141543P.
PR 02-JUL-1999; 99US-0142052P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 22-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 25-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 30-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 23-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 05-OCT-1999; 99US-0157117P.
PR 06-OCT-1999; 99US-0157753P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.

PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160768P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160772P.
 PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 25-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 26-OCT-1999; 99US-0161353P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 26-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161992P.
 PR 28-OCT-1999; 99US-0161993P.
 PR 29-OCT-1999; 99US-0162142P.

Query Match 68.6%; Score 1279.5; DB 3; Length 358;
 Best Local Similarity 67.0%; Pred. No. 6.7e-118;
 Matches 240; Conservative 39; Mismatches 72; Indels 7; Gaps 4;

QY 1 MAAKLFF---FLFLVLSALVALAGFEEDNPIRSVTORPDSIEPAAILGLVLSGCRHAFHFA 57
 DB 1 MSAKTILSSVVLVLAASAAANIGFDESPIRWSVDCGLREVESVSQILGQSRHVLSPA 60
 QY 58 RPARRYGKSYGSEBEIKRFGIFVENLAFIRSTNRKDLSTYTLGINQFADLTWEEFRNRL 117
 DB 61 RTHRYGKGYQNVBEMKLRFSIFKENLDLIRSTNKKGLSYKLVNQFADLTWQEFORTKL 120
 QY 118 GAAQNCSTAFHGNRFRVDGLPVTTRDREQIGVSPVKDQSGCS-WTFTTGALEAAATQ 176
 DB 121 GAAQNCSTATLKGSHKVTAAALPETKDWREDGIVSPVKDQSGCSWTFSTTGALEAAHQ 180
 QY 177 LTGS--TLSEQQLVDCASAFNNFGC--GGLPSQAFYVYKNGGIDTEQTYPLGVWGCNF 233
 DB 181 AFGKGISLSEQQLVDCAGAFNNYGCNGGLPSQAFYIKNSGGLDTERKAYPTGKDETCF 240
 QY 234 KOENVGVKVIDSINTITLGADELKHAAGLVLPVSPVAFVVKGFNLYKKGVSSTDCGRDP 293
 DB 241 SAENVGVQVLSNVNITLGADELKHAAGLVLPVSPVAFVVKGFNLYKKGVSSTDCGRDP 300
 QY 294 MDVNVHVLAVGVGVEDGTPYWLIRKNSWGTNGDNGYFKMELGKMGCVATCASYPIVA 351
 DB 301 MDVNVHVLAVGVGVEDGTPYWLIRKNSWGTNGDNGYFKMELGKMGCVATCASYPIVA 358

XX (HAYK/) HAVKIN-FRENKEL D.
 PA (PODS/) PODTOLSKI A.
 PA (DIXO/) DIXON R A.
 XX
 PI HAVKIN-Frenkel D, Podtolski A, Dixon RA;
 XX
 DR WPI; 2003-743794/70.
 XX
 PT Improving vanillin production in cultured Vanillin planifolia by
 PT supplementing tissue (e.g., embryo) culture of V. planifolia with malic
 PT acid or by subjecting the culture to heat or mechanical shear stress.
 XX
 PS Disclosure; Fig 6; 49pp; English.
 XX
 CC The present invention relates to a method for improving vanillin
 CC production in cultured Vanillin planifolia, and in intact plants. The
 CC method involves genetically engineering V. planifolia to overproduce
 CC enzymes associated with step(s) involved in vanillin biosynthesis in the
 CC plant. Transgenic cells and plants overproducing 4-hydroxybenzaldehyde
 CC synthase (4HBS) enzyme are also provided. The method is useful for
 CC improving vanillin production in cultured V. planifolia. The method
 CC results in the production of cultured cells or plants producing at least
 CC twice to 10 times more vanillin than cells or plants not produced by the
 CC new method. The present sequence represents a plant cysteine protease
 CC used for comparison to the V. planifolia 4HBS protein of the invention.
 XX
 SQ Sequence 360 AA;

Query Match 68.2%; Score 1272.5; DB 7; Length 360;
 Best Local Similarity 66.4%; Pred. No. 3.3e-117;
 Matches 239; Conservative 40; Mismatches 72; Indels 9; Gaps 4;

QY 1 MAAKLFF---FLFLVLSALVALAGFEEDNPIRSVTORPDSIEPAAILGLVLSGCRHAFHFA 57
 DB 1 MSAKTILSSVVLVLAASAAANIGFDESPIRWSVDCGLREVESVSQILGQSRHVLSPA 60
 QY 58 RPARRYGKSYGSEBEIKRFGIFVENLAFIRSTNRKDLSTYTLGINQFADLTWEEFRNRL 117
 DB 61 RTHRYGKGYQNVBEMKLRFSIFKENLDLIRSTNKKGLSYKLVNQFADLTWQEFORTKL 120
 QY 118 GAAQNCSTAFHGNRFRVDGLPVTTRDREQIGVSPVKDQSGCS-WTFTTGALEAAATQ 176
 DB 121 GAAQNCSTATLKGSHKVTAAALPETKDWREDGIVSPVKDQSGCSWTFSTTGALEAAHQ 180
 QY 177 L-----TGSTLSEQQLVDCASAFNNFGC--GGLPSQAFYVYKNGGIDTEQTYPLGVWGC 231
 DB 181 AFAFGKGISLSEQQLVDCAGAFNNYGCNGGLPSQAFYIKNSGGLDTERKAYPTGKDETC 240
 QY 232 NFKOENVGVKVIDSINTITLGADELKHAAGLVLPVSPVAFVVKGFNLYKKGVSSTDCGR 291
 DB 241 KFSANVGVQVLSNVNITLGADELKHAAGLVLPVSPVAFVVKGFNLYKKGVSSTDCGR 300
 QY 292 DPMOVNHAVALGVGVEDGTPYWLIRKNSWGTNGDNGYFKMELGKMGCVATCASYPIVA 351
 DB 301 TPDVNVHVLAVGVGVEDGTPYWLIRKNSWGTNGDNGYFKMELGKMGCVATCASYPIVA 360

RESULT 10
 ID AAW89557 standard; protein; 365 AA.
 AC AAW89557;
 XX
 DT 16-MAR-1999 (first entry)
 XX
 DE Triticum sp. cysteine proteinase #2.
 XX
 KW Triticum; wheat seed; cysteine proteinase; gluten; baking.
 OS Triticum sp.
 XX
 PN JP10327886-A.

PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160768P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160772P.
 PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 25-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 26-OCT-1999; 99US-0161353P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 26-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161992P.
 PR 28-OCT-1999; 99US-0161993P.
 PR 29-OCT-1999; 99US-0162142P.

Query Match 68.6%; Score 1279.5; DB 3; Length 358;
 Best Local Similarity 67.0%; Pred. No. 6.7e-118;
 Matches 240; Conservative 39; Mismatches 72; Indels 7; Gaps 4;

QY 1 MAAKLFF---FLFLVLSALVALAGFEEDNPIRSVTORPDSIEPAAILGLVLSGCRHAFHFA 57
 DB 1 MSAKTILSSVVLVLAASAAANIGFDESPIRWSVDCGLREVESVSQILGQSRHVLSPA 60
 QY 58 RPARRYGKSYGSEBEIKRFGIFVENLAFIRSTNRKDLSTYTLGINQFADLTWEEFRNRL 117
 DB 61 RTHRYGKGYQNVBEMKLRFSIFKENLDLIRSTNKKGLSYKLVNQFADLTWQEFORTKL 120
 QY 118 GAAQNCSTAFHGNRFRVDGLPVTTRDREQIGVSPVKDQSGCS-WTFTTGALEAAATQ 176
 DB 121 GAAQNCSTATLKGSHKVTAAALPETKDWREDGIVSPVKDQSGCSWTFSTTGALEAAHQ 180
 QY 177 LTGS--TLSEQQLVDCASAFNNFGC--GGLPSQAFYVYKNGGIDTEQTYPLGVWGCNF 233
 DB 181 AFGKGISLSEQQLVDCAGAFNNYGCNGGLPSQAFYIKNSGGLDTERKAYPTGKDETCF 240
 QY 234 KOENVGVKVIDSINTITLGADELKHAAGLVLPVSPVAFVVKGFNLYKKGVSSTDCGRDP 293
 DB 241 SAENVGVQVLSNVNITLGADELKHAAGLVLPVSPVAFVVKGFNLYKKGVSSTDCGRDP 300
 QY 294 MDVNVHVLAVGVGVEDGTPYWLIRKNSWGTNGDNGYFKMELGKMGCVATCASYPIVA 351
 DB 301 MDVNVHVLAVGVGVEDGTPYWLIRKNSWGTNGDNGYFKMELGKMGCVATCASYPIVA 358

RESULT 9
 ID ADC61395 standard; protein; 360 AA.
 AC ADC61395;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Arabidopsis thaliana cysteine protease.
 XX
 KW Vanillin production; plant; vanillin biosynthesis;
 KW 4-hydroxybenzaldehyde synthase; 4HBS; enzyme; cysteine protease.
 XX
 OS Arabidopsis thaliana.
 XX
 PN US2003070188-A1.
 XX
 PD 10-APR-2003.
 XX
 PF 28-FEB-2002; 2002US-00087714.
 XX
 PR 15-JUL-1997; 97US-0052604P.
 PR 15-JUL-1998; 98WO-US014895.
 PR 25-MAY-2000; 2000US-00462576.
 PR 28-FEB-2001; 2001US-0272415P.

```
XX 15-DEC-1998.
XX 27-MAR-1998; 98JP-00098140.
XX 31-MAR-1997; 97JP-00114946.
XX (SHOS) SHOWA SANGYO CO.
XX WPI; 1999-109255/10.
XX N-PSDB; AAV82457.
XX New DNA coding cysteine proteinase originating from wheat seed - useful
XX for improving gluten for use in bakery process.
XX Claim 2; Page 16-18; 29pp; Japanese.
XX The present sequence represents a cysteine proteinase isolated from wheat
XX seed (Triticum sp.). The cysteine proteinase is useful for improving
XX gluten for use in the bakery process
XX Sequence 365 AA;
XX
Query Match 68.1%; Score 1270; DB 2; Length 365;
Best Local Similarity 68.1%; Pred. No. 6e-117;
Matches 246; Conservative 36; Mismatches 67; Indels 12; Gaps 6;
QY 3 AKLLFFLLFLYSALSVALA-----GFEDNPISVTPRPS-TEPAILGLVLSGRHAFH 55
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 4 ARILLALALATATVAVASSSSSSSSFAVSNPIRPTVTERAASTLESTVLAALGRTRHALR 63
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 56 FARFARRYKSGSSEEEKKGFIEVENLAFIRSTNRKDLSTVLGINOFADLTWEEFRTN 115
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 64 FARFAVRVYKSVESAAYEMRFRIFESLSLEVRSTNRKGLSYRLGINFRFSDMSWEEFQAT 123
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 116 RLGAQNCSTAHGNHRFVD-GVLPVTRDREQGIIVSPVKDQSGGS-WTFSTTTGALEAA 173
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 124 RLGAQTCSATLGNHLNRDAALPETKDWREDGIVSPVKDQSGGSWTFSTTTGALEAA 183
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 174 YTLQFGS--TLSEQQLVDCASAFNFGC-GGLPSQAFYKVKNGIDTEQTPYLVGNGI 230
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 184 YTOATGKNISSEQQLVDCAGFNFGSGGLPSQAFYIKNGIDTEESYPYKGVNGV 243
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 231 CNFKQENGVKVIDSINITLGADELKHAAGLVLPVSVAFVKGFLNLYKGVYSSDTG 290
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 244 CHYKAENAVQVLDSVNTLNADELKNAAGLVLPVSVAFVKGFLNLYKGVYSSDTG 303
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 291 RDPMDVNHAVLAVGVGVEDGIPYLLIKNSGNTWGDNGYFKMELKKNMCGVATCASYPIV 350
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 304 TTPDDVNHAVLAVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGV 363
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 351 A 351
DB 364 A 364

RESULT 11
AAGS3786
ID AAGS3786 standard; protein; 358 AA.
AC AAGS3786;
XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 68511.
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-00301439.
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 03-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 29-MAR-1999; 99US-0126264P.
XX 01-APR-1999; 99US-0126785P.
XX 06-APR-1999; 99US-0127462P.
XX 08-APR-1999; 99US-0128234P.
XX 16-APR-1999; 99US-0128714P.
XX 19-APR-1999; 99US-0129845P.
XX 21-APR-1999; 99US-0130077P.
XX 23-APR-1999; 99US-0130449P.
XX 28-APR-1999; 99US-0130510P.
XX 30-APR-1999; 99US-0130891P.
XX 30-APR-1999; 99US-0131449P.
XX 30-APR-1999; 99US-0132048P.
XX 04-MAY-1999; 99US-0132407P.
XX 05-MAY-1999; 99US-0132484P.
XX 06-MAY-1999; 99US-0132485P.
XX 06-MAY-1999; 99US-0132486P.
XX 07-MAY-1999; 99US-0132487P.
XX 11-MAY-1999; 99US-0132863P.
XX 14-MAY-1999; 99US-0134218P.
XX 14-MAY-1999; 99US-0134219P.
XX 14-MAY-1999; 99US-0134221P.
XX 18-MAY-1999; 99US-0134370P.
XX 19-MAY-1999; 99US-0134376P.
XX 20-MAY-1999; 99US-0134941P.
XX 21-MAY-1999; 99US-0135124P.
XX 21-MAY-1999; 99US-0135353P.
XX 25-MAY-1999; 99US-0135629P.
XX 27-MAY-1999; 99US-0136021P.
XX 28-MAY-1999; 99US-0136392P.
XX 01-JUN-1999; 99US-0136782P.
XX 03-JUN-1999; 99US-0137222P.
XX 04-JUN-1999; 99US-0137502P.
XX 07-JUN-1999; 99US-0137724P.
XX 08-JUN-1999; 99US-0138094P.
XX 10-JUN-1999; 99US-0138540P.
XX 10-JUN-1999; 99US-0138847P.
XX 14-JUN-1999; 99US-0139119P.
XX 16-JUN-1999; 99US-0139453P.
XX 17-JUN-1999; 99US-0139492P.
XX 18-JUN-1999; 99US-0139454P.
XX 18-JUN-1999; 99US-0139455P.
XX 18-JUN-1999; 99US-0139456P.
XX 18-JUN-1999; 99US-0139457P.
XX 18-JUN-1999; 99US-0139458P.
XX 18-JUN-1999; 99US-0139459P.
XX 18-JUN-1999; 99US-0139460P.
XX 18-JUN-1999; 99US-0139461P.
XX 18-JUN-1999; 99US-0139462P.
XX 18-JUN-1999; 99US-0139463P.
XX 18-JUN-1999; 99US-0139750P.
XX 21-JUN-1999; 99US-0139763P.
XX 22-JUN-1999; 99US-0139817P.
XX 23-JUN-1999; 99US-0139899P.
XX 23-JUN-1999; 99US-0140353P.
XX 24-JUN-1999; 99US-0140354P.
XX 24-JUN-1999; 99US-0140695P.
XX 28-JUN-1999; 99US-0140823P.
XX 29-JUN-1999; 99US-0140981P.
XX 30-JUN-1999; 99US-0141287P.
XX 01-JUL-1999; 99US-0141842P.
XX 01-JUL-1999; 99US-0142154P.
```

PR	02-JUL-1999;	99US-0142055P.	PR	22-SEP-1999;	99US-0155139P.
PR	06-JUL-1999;	99US-0142390P.	PR	23-SEP-1999;	99US-0155486P.
PR	08-JUL-1999;	99US-0142803P.	PR	24-SEP-1999;	99US-0155653P.
PR	09-JUL-1999;	99US-0142920P.	PR	28-SEP-1999;	99US-0156458P.
PR	12-JUL-1999;	99US-0142977P.	PR	29-SEP-1999;	99US-0156596P.
PR	13-JUL-1999;	99US-0143544P.	PR	04-OCT-1999;	99US-0157117P.
PR	14-JUL-1999;	99US-0143624P.	PR	05-OCT-1999;	99US-0157753P.
PR	15-JUL-1999;	99US-0144005P.	PR	06-OCT-1999;	99US-0157865P.
PR	16-JUL-1999;	99US-0144086P.	PR	07-OCT-1999;	99US-0158029P.
PR	16-JUL-1999;	99US-0144086P.	PR	08-OCT-1999;	99US-0158232P.
PR	19-JUL-1999;	99US-0144325P.	PR	12-OCT-1999;	99US-0158363P.
PR	19-JUL-1999;	99US-0144331P.	PR	12-OCT-1999;	99US-0159293P.
PR	19-JUL-1999;	99US-0144332P.	PR	13-OCT-1999;	99US-0159294P.
PR	19-JUL-1999;	99US-0144333P.	PR	13-OCT-1999;	99US-0159295P.
PR	20-JUL-1999;	99US-0144352P.	PR	14-OCT-1999;	99US-0159329P.
PR	20-JUL-1999;	99US-0144633P.	PR	14-OCT-1999;	99US-0159330P.
PR	20-JUL-1999;	99US-0144633P.	PR	14-OCT-1999;	99US-0159331P.
PR	20-JUL-1999;	99US-0144884P.	PR	14-OCT-1999;	99US-0159637P.
PR	21-JUL-1999;	99US-0144814P.	PR	18-OCT-1999;	99US-0159638P.
PR	21-JUL-1999;	99US-0145086P.	PR	21-OCT-1999;	99US-0159584P.
PR	21-JUL-1999;	99US-0145088P.	PR	21-OCT-1999;	99US-0160767P.
PR	22-JUL-1999;	99US-0145085P.	PR	21-OCT-1999;	99US-0160768P.
PR	22-JUL-1999;	99US-0145087P.	PR	21-OCT-1999;	99US-0160770P.
PR	22-JUL-1999;	99US-0145192P.	PR	21-OCT-1999;	99US-0160814P.
PR	23-JUL-1999;	99US-0145145P.	PR	21-OCT-1999;	99US-0160815P.
PR	23-JUL-1999;	99US-0145218P.	PR	22-OCT-1999;	99US-0160980P.
PR	23-JUL-1999;	99US-0145224P.	PR	22-OCT-1999;	99US-0160981P.
PR	26-JUL-1999;	99US-0145276P.	PR	22-OCT-1999;	99US-0160989P.
PR	27-JUL-1999;	99US-0145913P.	PR	22-OCT-1999;	99US-0160981P.
PR	27-JUL-1999;	99US-0145918P.	PR	25-OCT-1999;	99US-0161404P.
PR	27-JUL-1999;	99US-0145919P.	PR	25-OCT-1999;	99US-0161405P.
PR	28-JUL-1999;	99US-0145951P.	PR	25-OCT-1999;	99US-0161406P.
PR	02-AUG-1999;	99US-0145386P.	PR	25-OCT-1999;	99US-0161359P.
PR	02-AUG-1999;	99US-0145388P.	PR	25-OCT-1999;	99US-0161360P.
PR	02-AUG-1999;	99US-0145389P.	PR	26-OCT-1999;	99US-0161361P.
PR	03-AUG-1999;	99US-0147038P.	PR	26-OCT-1999;	99US-0161920P.
PR	04-AUG-1999;	99US-0147204P.	PR	28-OCT-1999;	99US-0161922P.
PR	04-AUG-1999;	99US-0147302P.	PR	28-OCT-1999;	99US-0161933P.
PR	05-AUG-1999;	99US-0147192P.	PR	29-OCT-1999;	99US-0162142P.
PR	05-AUG-1999;	99US-0147260P.			
PR	06-AUG-1999;	99US-0147303P.			
PR	06-AUG-1999;	99US-0147416P.			
PR	06-AUG-1999;	99US-0147493P.			
PR	08-AUG-1999;	99US-0147935P.			
PR	09-AUG-1999;	99US-0148171P.			
PR	10-AUG-1999;	99US-0148319P.			
PR	11-AUG-1999;	99US-0148341P.			
PR	12-AUG-1999;	99US-0148565P.			
PR	13-AUG-1999;	99US-0148684P.			
PR	16-AUG-1999;	99US-01493368P.			
PR	17-AUG-1999;	99US-0149175P.			
PR	18-AUG-1999;	99US-0149426P.			
PR	20-AUG-1999;	99US-0149722P.			
PR	20-AUG-1999;	99US-0149723P.			
PR	20-AUG-1999;	99US-0149929P.			
PR	23-AUG-1999;	99US-0149902P.			
PR	23-AUG-1999;	99US-0149930P.			
PR	26-AUG-1999;	99US-0150566P.			
PR	26-AUG-1999;	99US-0150884P.			

Db 64 FAVRYGKSYSAEAEVRRRIIFSESLEEVSTNRKGLPYRLGGINRFSMSWEERFOATRLG 123
QY 119 AAQNCSATAGHNRFDV-GVLPVTRDREOGIVSPVKDQSGS-WTPESTTGAAEAAYTO 176
Db 124 AACTCSAT-KGNHLMEDAAALPETKDWREDGIVSPVKQAHCGSCWFTSTTGAAEAAYTO 182
QY 177 LTGS--TLSEQQLVDCASAFNFGC-GGLPSQAFYKVKINGGIDTEQTPYPLGVNGICNF 233
Db 183 ATGKNISLSEQQLVDCAGFNFGCGGLPSQAFYKVKINGGIDTEESYPYKGVNGVCHY 242
QY 234 KQNVGVKVIDSINITLGAEDELKNAVLVRPVSV-AFEVVKGENLYKKGVYSSDTCGRD 292
Db 243 KAENRAVOVLDSVNIILNAEDELKNAVLVRPVSVAAFQVIDGFRQYKSGVYTSDHGCTT 302
QY 293 PWDVNAVLAVGVVEDGIPYMLIKNSGCTNMGDNGYFKMELGKMKCVATCASYPIVA 351
Db 303 PWDVNAVLAVGVGVGVYVYVLIKNMGADMGDNGYFKMEMGKMKCAIATCASYPIVA 361

RESULT 14
AAG47119
ID AAG47119 standard; protein; 325 AA.
XX AC AAG47119;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 59354.
XX DE Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN BP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-00301439.
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0128264P.
PR 29-MAR-1999; 99US-0128785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130894P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134321P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0128264P.
PR 29-MAR-1999; 99US-0128785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130894P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134321P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 21-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140981P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145132P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.


```
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 65.5%; Score 1222; DB 3; Length 325;
Best Local Similarity 70.7%; Pred. No 2.9e-112;
Matches 224; Conservative 32; Mismatches 57; Indels 4; Gaps 3;

QY 39 IEPAILGVLSCHHAFARFARYKSYCSBEEIKKRFGIFVENLAFIRSTNRKDLST 98
Db 9 VEESVQILQSRHLSFARFTHRYGKYQNVVEEMKLRFSIFKENLDLIRSTNKKGLSYK 68

QY 99 LGINQFADLTWEFRNRLGAONCSATAGNHRFVDGLPVTRDREOGIVSPVKDQGS 158
Db 69 LGVNQFADLTWQFQTKLGAONCSATLKGSHKVTAAALPETKDREOGIVSPVKDQGG 128

QY 159 CGS-WTFSITGALAAAYTQITGS--TLSEQLVDCASAFNNFGC--GGLPQAFYVYKNG 214
Db 129 CGSCWTFSTTGALAAAYHQAQFGKIGISLSEQLVDCAGAFNNYGCNGGLPQAFYIKNSG 188

QY 215 GIDTEQTYPLGVNMGICNFKQENVGVKVIDSINITLGADELKHAHGLVRPVSVAFVYK 274
Db 189 GLDTEKAYPTYGKDETCKPSAENVGVQLNSVNITLGADELKHAHGLVRPVSVIAFEV 248

QY 275 GFNLKKGVSSTCGRDPMDVNHAVLAVGYGVVEDGIPYMLIKNSWGTNWDNGYFQML 334
Db 249 SFRLYKSGVYTDHSCGSTPMDVNHAVLAVGYGVVEDGVPYMLIKNSWGDWGDGKGYF 308

QY 335 GKNCMGVATCASYPIVA 351
Db 309 GKNCMGVATCASYPVVA 325
```

Search completed: March 29, 2004, 22:11:29
Job time : 89 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 29, 2004, 22:09:54 ; Search time 30 Seconds
(without alignments)
605.745 Million cell updates/sec

Title: US-10-087-714-2
Perfect score: 1865
Sequence: 1 MAKKLLPFLFLVLSVALS.....ELGKNMCGVATCASYPIVAV 352

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/FACTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/BACKFILES1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1300.5	69.7	378	4	US-09-325-932A-158
2	1252	67.1	361	3	US-09-120-365-75
3	1252	67.1	361	3	US-09-515-039-75
4	763	40.9	335	1	US-08-208-007A-10
5	763	40.9	335	1	US-08-330-121B-6
6	763	40.9	335	3	US-09-120-365-66
7	763	40.9	335	3	US-09-515-039-66
8	763	40.9	335	3	US-08-860-255A-6
9	763	40.9	335	4	US-08-915-095A-10
10	763	40.9	335	4	US-08-798-096-10
11	763	40.9	335	4	US-08-798-095A-10
12	763	40.9	335	4	US-09-953-956-10
13	763	40.9	335	4	US-08-553-125A-10
14	763	40.9	335	4	US-08-536-861-6
15	763	40.9	335	4	US-10-114-464-10
16	763	40.9	335	5	PCT-US95-13820-6
17	656.5	35.2	225	3	US-09-325-932A-162
18	656.5	35.2	225	3	US-09-120-365-67
19	656.5	35.2	225	3	US-09-515-039-67
20	603.5	32.4	339	3	US-09-120-365-62
21	603.5	32.4	339	3	US-09-120-365-80
22	603.5	32.4	339	3	US-09-515-039-80
23	603.5	32.4	339	3	US-08-515-039-80
24	594.5	31.9	323	3	US-09-120-365-61
25	594.5	31.9	323	3	US-09-515-039-61
26	594	31.8	500	4	US-09-325-932A-149
27	584.5	31.3	344	3	US-09-120-365-63

28	584.5	31.3	344	3	US-09-515-039-63	Sequence 63, Appl
29	584	31.3	386	4	US-09-325-932A-143	Sequence 143, Appl
30	583	31.3	374	3	US-08-821-994-67	Sequence 67, Appl
31	583	31.3	374	3	US-08-821-994-69	Sequence 69, Appl
32	581.5	31.2	467	3	US-08-821-994-83	Sequence 83, Appl
33	581	31.2	374	3	US-08-821-994-68	Sequence 68, Appl
34	580	31.1	370	4	US-09-325-932A-169	Sequence 169, Appl
35	573.5	30.8	443	3	US-08-821-994-65	Sequence 65, Appl
36	570	30.6	328	3	US-08-821-994-84	Sequence 84, Appl
37	562.5	30.2	351	1	US-09-500-651-2	Sequence 2, Appl
38	562.5	30.2	351	2	US-08-813-591-2	Sequence 2, Appl
39	562	30.1	374	3	US-08-821-994-70	Sequence 70, Appl
40	557	29.9	442	3	US-08-821-994-66	Sequence 66, Appl
41	556	29.8	380	3	US-09-120-365-76	Sequence 76, Appl
42	556	29.8	380	3	US-08-515-039-76	Sequence 76, Appl
43	556	29.8	380	3	US-08-860-255A-5	Sequence 5, Appl
44	554.5	29.7	457	3	US-09-120-365-72	Sequence 72, Appl
45	554.5	29.7	457	3	US-09-515-039-72	Sequence 72, Appl

ALIGNMENTS

RESULT 1
US-09-325-932A-158
; Sequence 158, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell death and their use in the modification of forestry plant level
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 158
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Eucalyptus grandis
US-09-325-932A-158

Query Match	69.7%	Score	1300.5	DB	4	Length	378
Best Local Similarity	70.4%	Pred. No.	8.7e-131				
Matches	247	Conservative	33	Mismatches	60	Indels	11
Gaps	5						
Qy	9	LLFLVLSALVALACGFEDNDPIRSVTQPD----	SIEPAILGLVLSGRHAFHFARFARYG	64			
Db	31	LLLVAVVVSAAASFESNPIRLP--PDGGLRDLESSIVQIVGRTTHAFSFARFARYG	87				
Qy	65	KSYGSBEIEKRGFGIFVENLAFIRSTNRKDLSTYTLGNOFADLTWEEFTRNLGAAQNC	124				
Db	88	KRYETABEIKLRPFIFRENKLIRSTNKKGLPYTLGVNKFADWSWEEFRHRLGAAQNC	147				
Qy	125	ATAGNHRFVDGVLPTVRDWRQGIIVSPVKDQSCGS-WTFSTTGCALEAAAYTQLTGS--T	181				
Db	148	ATTGKNHKLTDALPEMKDWREKIVSPDKQHGCGSWTFSTTGCALEAAVHQAFQKIS	207				
Qy	182	LSEQLVDCASAFNNFGC--GGLPSQAFYKYKNGIDTEQTYPYLVGMGICNFQENGV	240				
Db	208	LSEQLVDCAGAFNNFGCSGGLPSQAFYKYKNGIDTEEAAYPYTAVDGSCKFSADNVG	267				
Qy	241	KVIDSINITLGADELKHAIVLPVSAFVAVKGNLYKKGVYSSDTCCGRDPMVNHAV	300				
Db	268	QVLDSVINITLGADELKHAIVAFVPSVAFQVVDFFLYKSGVYTSITCGSTSDVNHAV	327				
Qy	301	LAVGVGEDGIPYWLIIKNSWGTWNGDNGYFMELGKNMCGVATCASYPIVA	351				
Db	328	LAVGVGEDGVFPLIIKNSWGDWDGHDGYFMENKMKMGVATCASYPVVA	378				

RESULT 2

US-09-120-365-75
; Sequence 75, Application US/09120365
; Patent No. 6103514

; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-14479
; CURRENT APPLICATION NUMBER: US/09/120,365
; CURRENT FILING DATE: 1998-07-22
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 75
; LENGTH: 361
; TYPE: PRT

; ORGANISM: Aleurain

US-09-120-365-75

Query Match 67.1%; Score 1252; DB 3; Length 361;
Best Local Similarity 67.3%; Pred. No. 1.3e-125;
Matches 241; Conservative 42; Mismatches 65; Indels 10; Gaps 7;

QY 3 AKLLFLLFLVSALSVALA---GFEEDNPIRSVTPQDPS--IEPAILGVLGSCRHAFHAF 58
DB 4 ARVLLALAVLATAAVALVASSSSFADSNPIRPTVTDRAASTLES AVLGA LGTRHALRFR 63
QY 59 FARYYKSGYSEEEIKKRFIFVENLAFIRSTNKRDLSTYLTGINQFADLTWEEPRTRNLG 118
DB 64 FAVRYKSGYESAAEVRFRFRIFSELSBEVRSNRRKGLPYRLGINRFSNDSWEEFQATRLG 123
QY 119 AAQNCSTAGHNRHFDV-GVLPVTRDWEQIGVSPVKDQSCGS-WTFSTTGALAAAYTQ 176
DB 124 AAQTCSTAGHNLMDAAALPETKDWREDDGIVSPVKNOAHCGSCWTFSTTGALAAAYTQ 183
QY 177 LTGS--TLSEQOLVDCASAFNNFGC-GGLPSQAFYVKYNGIGDTEQTYPLGVNMGICNF 233
DB 184 ATGKNISLSEQQLVDCAGFNFGCGGLPSQAFY-QYNGGIDTESYYPKGVNGVCHY 242
QY 234 KOENVGVKVIDSINTLTGADELKHAAGLVPRVPSVAPEVVKGNLYKKGVSSTDCGRDP 293
DB 243 KAENAAVQVLDVSNITLNAEDELKNAAGLVPRVPSVAQVIDGFRQYKSGVYTSDDHCGTTP 302
QY 294 MDVNHAVLAGVGVEDGIPYMLIKNSWGTNMGDNGYFKMELGKNMCGVATCASYPIVA 351
DB 303 DDVNHAVLAGVGVENGYPYMLTKNSWADGNDNGYFKMENGKNMCAIATCASYPVVA 360

RESULT 3

US-09-515-039-75
; Sequence 75, Application US/09515039
; Patent No. 6214599

; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-14479
; CURRENT APPLICATION NUMBER: US/09/515,039
; CURRENT FILING DATE: 2000-03-06
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 75
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Aleurain
US-09-515-039-75

Query Match 67.1%; Score 1252; DB 3; Length 361;
Best Local Similarity 67.3%; Pred. No. 1.3e-125;
Matches 241; Conservative 42; Mismatches 65; Indels 10; Gaps 7;

QY 3 AKLLFLLFLVSALSVALA---GFEEDNPIRSVTPQDPS--IEPAILGVLGSCRHAFHAF 58

DB 4 ARVLLALAVLATAAVALVASSSSFADSNPIRPTVTDRAASTLES AVLGA LGTRHALRFR 63
QY 59 FARYYKSGYSEEEIKKRFIFVENLAFIRSTNKRDLSTYLTGINQFADLTWEEPRTRNLG 118
DB 64 FAVRYKSGYESAAEVRFRFRIFSELSBEVRSNRRKGLPYRLGINRFSNDSWEEFQATRLG 123
QY 119 AAQNCSTAGHNRHFDV-GVLPVTRDWEQIGVSPVKDQSCGS-WTFSTTGALAAAYTQ 176
DB 124 AAQTCSTAGHNLMDAAALPETKDWREDDGIVSPVKNOAHCGSCWTFSTTGALAAAYTQ 183
QY 177 LTGS--TLSEQOLVDCASAFNNFGC-GGLPSQAFYVKYNGIGDTEQTYPLGVNMGICNF 233
DB 184 ATGKNISLSEQQLVDCAGFNFGCGGLPSQAFY-QYNGGIDTESYYPKGVNGVCHY 242
QY 234 KOENVGVKVIDSINTLTGADELKHAAGLVPRVPSVAPEVVKGNLYKKGVSSTDCGRDP 293
DB 243 KAENAAVQVLDVSNITLNAEDELKNAAGLVPRVPSVAQVIDGFRQYKSGVYTSDDHCGTTP 302
QY 294 MDVNHAVLAGVGVEDGIPYMLIKNSWGTNMGDNGYFKMELGKNMCGVATCASYPIVA 351
DB 303 DDVNHAVLAGVGVENGYPYMLTKNSWADGNDNGYFKMENGKNMCAIATCASYPVVA 360

RESULT 4

US-08-208-007A-10
; Sequence 10, Application US/08208007A
; Patent No. 5501969

; GENERAL INFORMATION:
; APPLICANT: HASTINGS, ET AL.
; TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSES: CARELLA, BYRNE, BAIN, GILFILLAN,
; CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068

COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/208,007A
; FILING DATE: March 8, 1994
; CLASSIFICATION: 435

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: No. 5501969e
; FILING DATE: No. 5501969e
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 335 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-08-208-007A-10

Query Match 40.9%; Score 763; DB 1; Length 335;
Best Local Similarity 52.1%; Pred. No. 3.1e-73;
Matches 159; Conservative 40; Mismatches 98; Indels 8; Gaps 7;

QY 54 FHFARFARYKSGYSEEEIKKRFIFVENLAFIRSTNKRDLSTYLTGINQFADLTWEEPR 113

```

Db      33  FHFKSWSKHKRTY-STEEYHRLQTFASNRKINAHNNNGHFFKMAALNQFSDMSFAEIK 91
QY      114  TNRL-GAAQNCATAHGNHRFVDGLVPTVRDREOG-IVSPVKDQSGCS-WTFSTTGAL 170
Db      92  HKYLWSEBPQNCAT-KSNYLRTGTGPPSPVDMRKGNFVSPVKNOGACGSCWTFSTTGAL 150
QY      171  EAAYTQLTGSLT--EQQLVDCASAFNFGC-GGLPSQAEYVYKNGGIDTEQTYPLGV 227
Db      151  ESAIAIATGKMLSLAEQQLVDCADQDFNNYCGQGLPSQAEYILYNKINGMEDTYPYQK 210
QY      228  MGICNFQKQENVGVKVIDSINITLGADELKHAHLVLRVPVSAFVVKGNLYKKGVISSD 287
Db      211  DGYCKFQPKAIGFVKDVANITTYDEAMVEAVALYNPVSFAFEVTDQDFMYRTGIYST 270
QY      288  TCGRDPMDVNHAVLAIVGYVEDGIPYWLKNSNGTNGDNGYFKMELGKMGCVATCASY 347
Db      271  SCHKTPDKVNHAVLAIVGYGKNGIPYWIKNWGPQMGNGYFLIERGKMGCLAACASY 330
QY      348  PIVAV 352
Db      331  PIVLV 335

```

RESULT 5

```

US-08-330-121B-6
; Sequence 6, Application US/08330121B
; Patent No. 5736357
; GENERAL INFORMATION:
; APPLICANT: Bromme, Dieter
; TITLE OF INVENTION: CATHEPSIN O2 PROTEASE
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton
; ADDRESSEE: & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 27-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-60261/DJB/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 335 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-330-121B-6

```

```

Query Match      40.9%; Score 763; DB 1; Length 335;
Best Local Similarity 52.1%; Pred. No. 3.1e-73;
Matches 159; Conservative 40; Mismatches 98; Indels 8; Gaps 7;

```

```

QY      54  FHFAFARYKSYGSEEEIKKRGIFVENLAFIRSTNRKDLSTYLGINQFADLTWEFR 113

```

```

Db      33  FHFKSWSKHKRTY-STEEYHRLQTFASNRKINAHNNNGHFFKMAALNQFSDMSFAEIK 91
QY      114  TNRL-GAAQNCATAHGNHRFVDGLVPTVRDREOG-IVSPVKDQSGCS-WTFSTTGAL 170
Db      92  HKYLWSEBPQNCAT-KSNYLRTGTGPPSPVDMRKGNFVSPVKNOGACGSCWTFSTTGAL 150
QY      171  EAAYTQLTGSLT--EQQLVDCASAFNFGC-GGLPSQAEYVYKNGGIDTEQTYPLGV 227
Db      151  ESAIAIATGKMLSLAEQQLVDCADQDFNNYCGQGLPSQAEYILYNKINGMEDTYPYQK 210
QY      228  MGICNFQKQENVGVKVIDSINITLGADELKHAHLVLRVPVSAFVVKGNLYKKGVISSD 287
Db      211  DGYCKFQPKAIGFVKDVANITTYDEAMVEAVALYNPVSFAFEVTDQDFMYRTGIYST 270
QY      288  TCGRDPMDVNHAVLAIVGYVEDGIPYWLKNSNGTNGDNGYFKMELGKMGCVATCASY 347
Db      271  SCHKTPDKVNHAVLAIVGYGKNGIPYWIKNWGPQMGNGYFLIERGKMGCLAACASY 330
QY      348  PIVAV 352
Db      331  PIVLV 335

```

RESULT 6

```

US-09-120-365-66
; Sequence 66, Application US/09120365
; Patent No. 6103514
; GENERAL INFORMATION:
; APPLICANT: Natori, Shunji
; TITLE OF INVENTION: NEW PROTEASE
; FILE REFERENCE: 32290-144749
; CURRENT APPLICATION NUMBER: US/09/120,365
; CURRENT FILING DATE: 1998-07-22
; EARLIER APPLICATION NUMBER: JP 9-333 474
; EARLIER FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 66
; TYPE: PRT
; LENGTH: 335
; ORGANISM: Homo sapien cathepsin H
US-09-120-365-66

```

```

Query Match      40.9%; Score 763; DB 3; Length 335;
Best Local Similarity 52.1%; Pred. No. 3.1e-73;
Matches 159; Conservative 40; Mismatches 98; Indels 8; Gaps 7;

```

```

QY      54  FHFAFARYKSYGSEEEIKKRGIFVENLAFIRSTNRKDLSTYLGINQFADLTWEFR 113
Db      33  FHFKSWSKHKRTY-STEEYHRLQTFASNRKINAHNNNGHFFKMAALNQFSDMSFAEIK 91
QY      114  TNRL-GAAQNCATAHGNHRFVDGLVPTVRDREOG-IVSPVKDQSGCS-WTFSTTGAL 170
Db      92  HKYLWSEBPQNCAT-KSNYLRTGTGPPSPVDMRKGNFVSPVKNOGACGSCWTFSTTGAL 150
QY      171  EAAYTQLTGSLT--EQQLVDCASAFNFGC-GGLPSQAEYVYKNGGIDTEQTYPLGV 227
Db      151  ESAIAIATGKMLSLAEQQLVDCADQDFNNYCGQGLPSQAEYILYNKINGMEDTYPYQK 210
QY      228  MGICNFQKQENVGVKVIDSINITLGADELKHAHLVLRVPVSAFVVKGNLYKKGVISSD 287
Db      211  DGYCKFQPKAIGFVKDVANITTYDEAMVEAVALYNPVSFAFEVTDQDFMYRTGIYST 270
QY      288  TCGRDPMDVNHAVLAIVGYVEDGIPYWLKNSNGTNGDNGYFKMELGKMGCVATCASY 347
Db      271  SCHKTPDKVNHAVLAIVGYGKNGIPYWIKNWGPQMGNGYFLIERGKMGCLAACASY 330
QY      348  PIVAV 352
Db      331  PIVLV 335

```

RESULT 7


```

Db      33  FHFKSMWMSKHKRTY-STEEYHRLQTFASNWRKINAHNNGNHTFKVALNQFSDMSFAEIK 91
QY      114  TNRL-GAAQNCATAGHNRHFDVGLPVTRDWEQ-IVSPVKDQSCGS-WTFSTTGAL 170
Db      92  HKYLWSEPQNCAT-KSNYLRTGTPPSPVDWRKGNFVSPVKNQAGCGSCWTFSTTGAL 150
QY      171  EAAVTQLTGSTLS--EQQLVDCASAFNNFC--GGLPSQAFYVKYNGGIDTEQTPYLVG 227
Db      151  ESAIAIATGKMLSLAEQQLVDCADQFNQYCGQGLPSQAFYILYKNGIMGEDTYPYQKG 210
QY      228  MGINCFKQENGVKVIDSINITLGADELKHAVLVRPVSVAFEVVKVFNLYKKGVSDD 287
Db      211  DGYCKFQPKAIGFVKDVANITIIDEEAMVEALYNPVSFAFEVTQDFMMYRTGIYST 270
QY      288  TCGRDPMDVNHAVLAVGYVEDGIPYWLKNSGNTWGDNGYFKMELKNNMCGVATCASY 347
Db      271  SCHKTPDKVNHAVLAVGYGKNGIPYWIKNWGPQWGMNGYFLIERGKNMCGLAACASY 330
QY      348  PIVAV 352
Db      331  PIVLV 335

```

```

RESULT 10
US-08-798-096-10
; Sequence 10, Application US/08798096
; Patent No. 6387682
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D2
; CURRENT APPLICATION NUMBER: US/08/798,096
; PRIOR FILING DATE: 1997-02-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 10
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-798-096-10

```

```

Query Match      40.9%; Score 763; DB 4; Length 335;
Best Local Similarity 52.1%; Pred. No. 3.1e-73;
Matches 159; Conservative 40; Mismatches 98; Indels 8; Gaps 7;

QY      54  FHFAFPARYKSGYSGSEEEKKRGIFVENLAFIRSTNRKDLSTYLGINQFADLTWEPR 113
Db      33  FHFKSMWMSKHKRTY-STEEYHRLQTFASNWRKINAHNNGNHTFKVALNQFSDMSFAEIK 91
QY      114  TNRL-GAAQNCATAGHNRHFDVGLPVTRDWEQ-IVSPVKDQSCGS-WTFSTTGAL 170
Db      92  HKYLWSEPQNCAT-KSNYLRTGTPPSPVDWRKGNFVSPVKNQAGCGSCWTFSTTGAL 150
QY      171  EAAVTQLTGSTLS--EQQLVDCASAFNNFC--GGLPSQAFYVKYNGGIDTEQTPYLVG 227
Db      151  ESAIAIATGKMLSLAEQQLVDCADQFNQYCGQGLPSQAFYILYKNGIMGEDTYPYQKG 210
QY      228  MGINCFKQENGVKVIDSINITLGADELKHAVLVRPVSVAFEVVKVFNLYKKGVSDD 287
Db      211  DGYCKFQPKAIGFVKDVANITIIDEEAMVEALYNPVSFAFEVTQDFMMYRTGIYST 270
QY      288  TCGRDPMDVNHAVLAVGYVEDGIPYWLKNSGNTWGDNGYFKMELKNNMCGVATCASY 347
Db      271  SCHKTPDKVNHAVLAVGYGKNGIPYWIKNWGPQWGMNGYFLIERGKNMCGLAACASY 330
QY      348  PIVAV 352
Db      331  PIVLV 335

```

```

Query Match      40.9%; Score 763; DB 4; Length 335;
Best Local Similarity 52.1%; Pred. No. 3.1e-73;
Matches 159; Conservative 40; Mismatches 98; Indels 8; Gaps 7;

```

```

QY      54  FHFAFPARYKSGYSGSEEEKKRGIFVENLAFIRSTNRKDLSTYLGINQFADLTWEPR 113
Db      33  FHFKSMWMSKHKRTY-STEEYHRLQTFASNWRKINAHNNGNHTFKVALNQFSDMSFAEIK 91
QY      114  TNRL-GAAQNCATAGHNRHFDVGLPVTRDWEQ-IVSPVKDQSCGS-WTFSTTGAL 170
Db      92  HKYLWSEPQNCAT-KSNYLRTGTPPSPVDWRKGNFVSPVKNQAGCGSCWTFSTTGAL 150
QY      171  EAAVTQLTGSTLS--EQQLVDCASAFNNFC--GGLPSQAFYVKYNGGIDTEQTPYLVG 227
Db      151  ESAIAIATGKMLSLAEQQLVDCADQFNQYCGQGLPSQAFYILYKNGIMGEDTYPYQKG 210
QY      228  MGINCFKQENGVKVIDSINITLGADELKHAVLVRPVSVAFEVVKVFNLYKKGVSDD 287
Db      211  DGYCKFQPKAIGFVKDVANITIIDEEAMVEALYNPVSFAFEVTQDFMMYRTGIYST 270
QY      288  TCGRDPMDVNHAVLAVGYVEDGIPYWLKNSGNTWGDNGYFKMELKNNMCGVATCASY 347
Db      271  SCHKTPDKVNHAVLAVGYGKNGIPYWIKNWGPQWGMNGYFLIERGKNMCGLAACASY 330
QY      348  PIVAV 352
Db      331  PIVLV 335

```

```

RESULT 11
US-08-798-095A-10

```

```

; Sequence 10, Application US/08798095A
; Patent No. 6423507
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D3
; CURRENT APPLICATION NUMBER: US/08/798,095A
; PRIOR FILING DATE: 1997-02-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 10
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-798-095A-10

```

```

Query Match      40.9%; Score 763; DB 4; Length 335;
Best Local Similarity 52.1%; Pred. No. 3.1e-73;
Matches 159; Conservative 40; Mismatches 98; Indels 8; Gaps 7;

```

```

QY      54  FHFAFPARYKSGYSGSEEEKKRGIFVENLAFIRSTNRKDLSTYLGINQFADLTWEPR 113
Db      33  FHFKSMWMSKHKRTY-STEEYHRLQTFASNWRKINAHNNGNHTFKVALNQFSDMSFAEIK 91
QY      114  TNRL-GAAQNCATAGHNRHFDVGLPVTRDWEQ-IVSPVKDQSCGS-WTFSTTGAL 170
Db      92  HKYLWSEPQNCAT-KSNYLRTGTPPSPVDWRKGNFVSPVKNQAGCGSCWTFSTTGAL 150
QY      171  EAAVTQLTGSTLS--EQQLVDCASAFNNFC--GGLPSQAFYVKYNGGIDTEQTPYLVG 227
Db      151  ESAIAIATGKMLSLAEQQLVDCADQFNQYCGQGLPSQAFYILYKNGIMGEDTYPYQKG 210
QY      228  MGINCFKQENGVKVIDSINITLGADELKHAVLVRPVSVAFEVVKVFNLYKKGVSDD 287
Db      211  DGYCKFQPKAIGFVKDVANITIIDEEAMVEALYNPVSFAFEVTQDFMMYRTGIYST 270
QY      288  TCGRDPMDVNHAVLAVGYVEDGIPYWLKNSGNTWGDNGYFKMELKNNMCGVATCASY 347
Db      271  SCHKTPDKVNHAVLAVGYGKNGIPYWIKNWGPQWGMNGYFLIERGKNMCGLAACASY 330
QY      348  PIVAV 352
Db      331  PIVLV 335

```

```

RESULT 12
US-09-953-956-10
; Sequence 10, Application US/09953956
; Patent No. 6475487
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D2D1
; CURRENT APPLICATION NUMBER: US/09/953,956
; CURRENT FILING DATE: 2001-09-18
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 10
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-953-956-10

```

```

Query Match      40.9%; Score 763; DB 4; Length 335;
Best Local Similarity 52.1%; Pred. No. 3.1e-73;
Matches 159; Conservative 40; Mismatches 98; Indels 8; Gaps 7;

QY      54  FHFAFPARYKSGYSGSEEEKKRGIFVENLAFIRSTNRKDLSTYLGINQFADLTWEPR 113
Db      33  FHFKSMWMSKHKRTY-STEEYHRLQTFASNWRKINAHNNGNHTFKVALNQFSDMSFAEIK 91

```

114 TNRL-GAAQNCATAGHNRHFRVDGLPVTDRDREQ-IVSPVKDQSCGS-WTFSTTGAL 170
92 HKYLWSEPQNCAT-KSNYLRTGTPPPSVDRKKNFVSPVKQACGSCWTFSTTGAL 150
171 EAAATQLTGSLT--EQQLVDCASAFNFGC-GGLPSQAFVYKNGGIDTQTPYLV 227
151 ESAIATATGKMLSLAEQQLVDCADQDFNNYCGGLPSQAFVYLYNKGIMGSDTYPYQK 210
228 MGICNFKQENVGVKVIDSINITLGADELKHAVALVPRVPSVAPEVVKGNLYKGVYSSD 287
211 DGCKFQPKAIGFVKDVANITYDEAMVEALYNFVSFAFVTDQFMVYRTGIYST 270
288 TCGRDPMDVNHAVLAGVGVEDGIPYWLKNSGNTWNGDGYFKMELGKMGCVATCASY 347
271 SCHKTPDKVNHAVLAGVGEKNGIPYWKNSGQWNGYFLIERGKNMCGLAACASY 330
348 PIVAV 352
331 PIPLV 335

RESULT 13
US-08-533-125A-10
; Sequence 10, Application US/08533125A
; Patent No. 6475766
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: P107D1
; CURRENT APPLICATION NUMBER: US/08/533.125A
; CURRENT FILING DATE: 1995-11-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 10
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-533-125A-10

Query Match 40.9%; Score 763; DB 4; Length 335;
Best Local Similarity 52.1%; Pred. No. 3.1e-73;
Matches 159; Conservative 40; Mismatches 98; Indels 8; Gaps 7;
QY 54 PHEAFRARRYKSYGSEEEIKRFGIFVENLAFIRSTNRKDLSTLGINOPADLTWEEFR 113
DB 33 FHKSWMSKHKRTY-STEEYHRLQTFASNRKINAHNNNGHTFMALNQFSDMSFAEIK 91
QY 114 TNRL-GAAQNCATAGHNRHFRVDGLPVTDRDREQ-IVSPVKDQSCGS-WTFSTTGAL 170
DB 92 HKYLWSEPQNCAT-KSNYLRTGTPPPSVDRKKNFVSPVKQACGSCWTFSTTGAL 150
QY 171 EAAATQLTGSLT--EQQLVDCASAFNFGC-GGLPSQAFVYKNGGIDTQTPYLV 227
DB 151 ESAIATATGKMLSLAEQQLVDCADQDFNNYCGGLPSQAFVYLYNKGIMGSDTYPYQK 210
QY 228 MGICNFKQENVGVKVIDSINITLGADELKHAVALVPRVPSVAPEVVKGNLYKGVYSSD 287
DB 211 DGCKFQPKAIGFVKDVANITYDEAMVEALYNFVSFAFVTDQFMVYRTGIYST 270
QY 288 TCGRDPMDVNHAVLAGVGVEDGIPYWLKNSGNTWNGDGYFKMELGKMGCVATCASY 347
DB 271 SCHKTPDKVNHAVLAGVGEKNGIPYWKNSGQWNGYFLIERGKNMCGLAACASY 330
QY 348 PIVAV 352
DB 331 PIPLV 335

RESULT 14
US-08-536-861-6
; Sequence 6, Application US/08536861
; Patent No. 6544767
; GENERAL INFORMATION:

APPLICANT: Bromme, Dieter
APPLICANT: Okamoto, Kathleen
TITLE OF INVENTION: CATHEPSIN O2 PROTEASE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/536.861
FILING DATE: 02-OCT-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-60261-1/DJB/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 335 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-536-861-6

Query Match 40.9%; Score 763; DB 4; Length 335;
Best Local Similarity 52.1%; Pred. No. 3.1e-73;
Matches 159; Conservative 40; Mismatches 98; Indels 8; Gaps 7;
QY 54 PHEAFRARRYKSYGSEEEIKRFGIFVENLAFIRSTNRKDLSTLGINOPADLTWEEFR 113
DB 33 FHKSWMSKHKRTY-STEEYHRLQTFASNRKINAHNNNGHTFMALNQFSDMSFAEIK 91
QY 114 TNRL-GAAQNCATAGHNRHFRVDGLPVTDRDREQ-IVSPVKDQSCGS-WTFSTTGAL 170
DB 92 HKYLWSEPQNCAT-KSNYLRTGTPPPSVDRKKNFVSPVKQACGSCWTFSTTGAL 150
QY 171 EAAATQLTGSLT--EQQLVDCASAFNFGC-GGLPSQAFVYKNGGIDTQTPYLV 227
DB 151 ESAIATATGKMLSLAEQQLVDCADQDFNNYCGGLPSQAFVYLYNKGIMGSDTYPYQK 210
QY 228 MGICNFKQENVGVKVIDSINITLGADELKHAVALVPRVPSVAPEVVKGNLYKGVYSSD 287
DB 211 DGCKFQPKAIGFVKDVANITYDEAMVEALYNFVSFAFVTDQFMVYRTGIYST 270
QY 288 TCGRDPMDVNHAVLAGVGVEDGIPYWLKNSGNTWNGDGYFKMELGKMGCVATCASY 347
DB 271 SCHKTPDKVNHAVLAGVGEKNGIPYWKNSGQWNGYFLIERGKNMCGLAACASY 330
QY 348 PIVAV 352
DB 331 PIPLV 335

RESULT 15
US-10-114-464-10
; Sequence 10, Application US/10114464
; Patent No. 6680375
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN

```
FILE REFERENCE: PF107DS
CURRENT APPLICATION NUMBER: US/10/114,464
CURRENT FILING DATE: 2002-04-03
PRIOR APPLICATION NUMBER: 08/553,125
PRIOR FILING DATE: 1995-11-07
PRIOR APPLICATION NUMBER: 08/208,007
PRIOR FILING DATE: 1994-03-08
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 335
TYPE: PRT
ORGANISM: Homo sapiens
US-10-114-464-10

Query Match      40.9%; Score 763; DB 4; Length 335;
Best Local Similarity 52.1%; Pred. No. 3.1e-73;
Matches 159; Conservative 40; Mismatches 98; Indels 8; Gaps 7;

QY  54 FHFAFARYGKSYGSEEEIKKGFIVENLAFIRSTNRKDLSTYTLGINOFADLTWEEPR 113
DB  33 FHFKSWMSKHKTY-STEEYHRLQTFASNWRKINAHNNNGHTFKNALNQFSDMSFAEIK 91

QY  114 TNRL-GAAQNCSATAGNHRFVDGLPVTRDWEQG-IVSPVKDQSGCS-WTFSTTGAL 170
DB  92 HKYLNSEPCNSAT-KSNYLRCGTGYPSPVDWRKKGNFVSPVKQAGCGSCWTFSTTGAL 150

QY  171 EAYTQLTGSLTSL--EQQLVDCASAFNFGC--GGLPSQAFHYVKNKGIDTQIYPLGV 227
DB  151 ESAIAIATGKMLSLABQQQLVDCADQDFNNYGCQGLPSQAFHYLYLNKIMGEDTYPYQOK 210

QY  228 MGICNFKQENVGVKVIDSINITLGADELKHAVGLVRPVSAFVGVVKNLYKGVSSD 287
DB  211 DGYCKFQPKAIGFYKDVANITIIYDEEAMVEAVALYNPVSAFVTDQFMRYRGIYST 270

QY  288 TCGRDPMDYNHVLAVGYGVEDGIPYWLKNSWGTNWDNGYFKMELGNMCGVATCASY 347
DB  271 SCHKTPDKYNHVLAVGYGKNGIPYWIIVKNSWGPQWGMNGYFLIERGKNMCGLAACASY 330

QY  348 PIVAV 352
DB  331 PIPLV 335
```

Search completed: March 29, 2004, 22:16:57
Job time : 31 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 29, 2004, 22:15:19 ; Search time 332 Seconds
(without alignments)
277.425 Million cell updates/sec

Title: US-10-087-714-2

Perfect score: 1865

Sequence: 1 MAAKLFFLLFLVSALVAL.....ELGKNMGVATCASYPIVAV 352

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1065169 seqs, 261661801 residues

Total number of hits satisfying chosen parameters: 1065169

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1865	100.0	352	14	US-10-087-714-2
2	1323	70.9	363	14	Sequence 2, Appli
3	1321.5	70.9	368	12	Sequence 24, Appl
4	1320.5	70.8	368	12	Sequence 56425, A
5	1319.5	70.8	368	12	Sequence 59088, A
6	1319.5	70.8	368	12	Sequence 42679, A
7	1310.5	70.3	340	12	Sequence 69552, A
8	1302.5	69.8	360	14	Sequence 58728, A
9	1300.5	69.7	378	14	Sequence 23, Appl
10	1286	69.0	360	14	Sequence 158, App
11	1280	68.6	353	12	Sequence 195993, A
12	1280	68.6	353	12	Sequence 44426, A
13	1272.5	68.2	360	14	Sequence 20, Appl
14	1261.5	67.6	362	14	Sequence 25, Appl
15	1255.5	67.3	362	14	Sequence 21, Appl

16	1252	67.1	308	12	US-10-425-114-51874
17	1249	67.0	361	12	Sequence 51874, A
18	1236.5	66.3	355	12	Sequence 49461, A
19	1148	61.6	357	14	Sequence 283471, A
20	852.5	45.7	193	12	Sequence 283, App
21	845.5	45.3	277	14	Sequence 46437, A
22	838	44.9	201	12	Sequence 280, App
23	828.5	44.4	185	12	Sequence 42955, A
24	780.5	41.8	168	12	Sequence 57524, A
25	763	40.9	335	9	Sequence 46575, A
26	763	40.9	335	13	Sequence 10, Appl
27	763	40.9	335	14	Sequence 10, Appl
28	759	40.7	364	9	Sequence 478, Appl
29	656.5	35.2	225	14	Sequence 162, App
30	609	32.7	470	14	Sequence 278, App
31	598.5	32.1	381	12	Sequence 43777, A
32	594.5	31.9	452	14	Sequence 276, App
33	594	31.8	500	14	Sequence 143, App
34	592.5	31.8	497	12	Sequence 65121, A
35	590.5	31.7	494	12	Sequence 49414, A
36	590.5	31.7	495	12	Sequence 67151, A
37	590.5	31.7	495	12	Sequence 67152, A
38	589	31.6	464	15	Sequence 7, Appli
39	589	31.6	479	12	Sequence 4438, A
40	589	31.6	479	12	Sequence 55418, A
41	584	31.3	386	14	Sequence 143, App
42	584	31.3	478	12	Sequence 44927, A
43	584	31.3	478	12	Sequence 51948, A
44	584	31.3	478	12	Sequence 53490, A
45	583.5	31.3	480	12	Sequence 47349, A

ALIGNMENTS

RESULT 1

US-10-087-714-2
; Sequence 2, Application US/10087714
; Publication No. US20030070188A1
; GENERAL INFORMATION:
; APPLICANT: Havkin-Frenkel, Daphna
; APPLICANT: Podstolski, Andrzej
; APPLICANT: Dixon, Richard A.
; TITLE OF INVENTION: Vanillin Biosynthetic Pathway Enzyme From Vanilla Planifolia
; FILE REFERENCE: DMC10099
; CURRENT APPLICATION NUMBER: US/10/087,714
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: 09/462,576
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: PCT/US98/14895
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: 60/052,604
; PRIOR FILING DATE: 1997-07-15
; PRIOR APPLICATION NUMBER: 60/272,415
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent version 3.1
; SEQ ID NO 2
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Vanilla planifolia
US-10-087-714-2

Query Match	100.0%	Score 1865	DB 14	Length 352
Best Local Similarity	100.0%	Pred. No. 3.5e+188		
Matches 352	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	1	MAAKLFFLLFLVSALVALAGFEEDNPIRSVTPQSDIEPAILGVLSGRHAFHAFPA	60	
Db	1	MAAKLFFLLFLVSALVALAGFEEDNPIRSVTPQSDIEPAILGVLSGRHAFHAFPA	60	
Qy	61	RRYKSGVSEETKKEFGIVENLAFIRSNKDLSTVLGINQFADLTWEEPTNRLGAA	120	

Db 61 RRYKSYGSEETIKKREGIFVENLAFIRSTNRKDLSTYTLGINOFADLTWEEFRTNRLGAA 120
QY 121 QNCSATAGHNRHFDVGLPVTRDREOGIVSPVKDQSGSWTFTTGALEAAAYTOLTGS 180
Db 121 QNCSATAGHNRHFDVGLPVTRDREOGIVSPVKDQSGSWTFTTGALEAAAYTOLTGS 180
QY 181 TLSEQQQLVDCASAFNNFGCGGLPSQAFYKYKNGGIDTEQTYPLGVGMGICNFKQENGV 240
Db 181 TLSEQQQLVDCASAFNNFGCGGLPSQAFYKYKNGGIDTEQTYPLGVGMGICNFKQENGV 240
QY 241 KVIDSINITILGADELKHAVLRPVSVAFVVKGNLYKKGVYSSDTCGRDMDVNHAV 300
Db 241 KVIDSINITILGADELKHAVLRPVSVAFVVKGNLYKKGVYSSDTCGRDMDVNHAV 300
QY 301 LAVGYGVEDGIPYWLKNSWGTNGDNGYFKMELGKMGCVATCASYPVAV 352
Db 301 LAVGYGVEDGIPYWLKNSWGTNGDNGYFKMELGKMGCVATCASYPVAV 352

RESULT 2
US-10-087-714-24
; Sequence 24, Application US/10087714
; Publication No. US20030070188A1
; GENERAL INFORMATION:
; APPLICANT: Havkin-Frenkel, Daphna
; APPLICANT: Podstolski, Andrzej
; APPLICANT: Dixon, Richard A.
; TITLE OF INVENTION: Vanillin Biosynthetic Pathway Enzyme From Vanilla Planifolia
; CURRENT APPLICATION NUMBER: US/10/087,714
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: 09/462,576
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: PCT/US98/14895
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: 60/052,604
; PRIOR FILING DATE: 1997-07-15
; PRIOR APPLICATION NUMBER: 60/272,415
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 24
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Zea mays
US-10-087-714-24

Query Match 70.9%; Score 1323; DB 14; Length 363;
Best Local Similarity 70.8%; Pred. No. 7.2e-131;
Matches 257; Conservative 30; Mismatches 64; Indels 12; Gaps 6;

QY 1 MAAKLLPFLFL----VSALSVALAGFEEDNPISVTQRPDS-IEPAILGVLSGCRHAF 54
Db 1 MAHRRIILLVAAVAATSAVAASGDDSDNPPIRPTVTDRAASALESTVFAALGRTRDAL 60
QY 55 HPARFARYGKSYGSEETIKKREGIFVENLAFIRSTNRKDLSTYTLGINOFADLTWEEFRT 114
Db 61 RFARFARYGKSYGSEETIKKREGIFVENLAFIRSTNRKDLSTYTLGINOFADLTWEEFRT 120
QY 115 NRLGAQNCSTAHGNHRF--VDGVLVTRDREOGIVSPVKDQSGCS-WTFSTTGALE 171
Db 121 TRLGAQNCSTALTCGNHRMAAAVALPETKDWREDGIVSPVKDQSGCSWTFSTTGALE 180
QY 172 AAYTOLTGS--TLSEQQQLVDCASAFNNFGC--GGLPSQAFYKYKNGGIDTEQTYPLGVGM 228
Db 181 AAYTQATKPISSLSEQQQLVDCGFAFNNFGCGGLPSQAFYKYKNGGLDTEESYFYQGVN 240
QY 229 GICNFKQENGVKVIDSINITILGADELKHAVLRPVSVAFVVKGNLYKKGVYSSDTCGR 288
Db 241 GICNFKQENGVKVIDSINITILGADELKHAVLRPVSVAFVVKGNLYKKGVYSSDTCGR 300
QY 289 CGRDPMDVNHAVLAVGYGVEDGIPYWLKNSWGTNGDNGYFKMELGKMGCVATCASYP 348

Db 301 CGTTPMDVNHAVLAVGYGVEDGIPYWLKNSWGTNGDNGYFKMELGKMGCVATCASYP 360
QY 349 IVA 351
Db 361 IVA 363

RESULT 3
US-10-425-114-56425
; Sequence 56425, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 56425
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB189-021-F2_FLI.pep
US-10-425-114-56425

Query Match 70.9%; Score 1321.5; DB 12; Length 368;
Best Local Similarity 71.7%; Pred. No. 1.1e-130;
Matches 258; Conservative 27; Mismatches 66; Indels 9; Gaps 6;

QY 1 MAAKLLPFLFLVLSALSVALA--GFEEDNPISVTQRPDS-IEPAILGVLSGCRHAFHA 57
Db 9 MAPRELLVLAVALAATAAANSAGPADSNPIRPTVTDRAASALESTVFAALGRTRDALRFA 68
QY 58 REARFARYGKSYGSEETIKKREGIFVENLAFIRSTNRKDLSTYTLGINOFADLTWEEFRTNRL 117
Db 69 RFARFARYGKSYGSEETIKKREGIFVENLAFIRSTNRKDLSTYTLGINOFADLTWEEFRTNRL 128
QY 118 GAAQNCSTAHGNHRF--VDGVLVTRDREOGIVSPVKDQSGCS-WTFSTTGALEAA 174
Db 129 GAAQNCSTALTCGNHRMAAAVALPETKDWREDGIVSPVKDQSGCSWTFSTTGALEAA 188
QY 175 TOLTGS--TLSEQQQLVDCASAFNNFGC--GGLPSQAFYKYKNGGIDTEQTYPLGVGMGIC 231
Db 189 TQATKPISSLSEQQQLVDCGFAFNNFGCGGLPSQAFYKYKNGGLDTEESYFYQGVN 248
QY 232 NFKQENGVKVIDSINITILGADELKHAVLRPVSVAFVVKGNLYKKGVYSSDTCGR 291
Db 249 KFKENGVKVIDSINITILGADELKHAVLRPVSVAFVVKGNLYKKGVYSSDTCGR 308
QY 292 DMDVNHAVLAVGYGVEDGIPYWLKNSWGTNGDNGYFKMELGKMGCVATCASYPVAV 351
Db 309 TMDVNHAVLAVGYGVEDGIPYWLKNSWGTNGDNGYFKMELGKMGCVATCASYPVAV 368

RESULT 4
US-10-425-114-59088
; Sequence 59088, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With


```

Db      88 KYETAEKILPEIFRENKLIIRSTNKKGLTYITGVNKFADSWEBFRHRLGAQNC 147
Qy      125 ATAHGNHREVDGVLPTVRDREQGIIVSPVKDQSCGS-WTFSTTGALAAAYTQLTGS--T 181
Db      148 ATTGNHKLTDALPEMDKREKIGVSPVKDQCHGCGSCWTFTSTTGALAAAYHQAQKQIS 207
Qy      182 LSEQQLVDCASAFNNFGC--GLPSQAFVYKNGGIDTEQTPYLVGMGICNFKQENYGV 240
Db      208 LSEQQLVDCAGAFNFGCGGLPSQAFVYKNGGLDTEAYPTAVDQCKFSADNVGV 267
Qy      241 KVIDSINITLGADELKHAVGLVRPVSVAFEVVKGNLYKKGVSDDTCGRDPMVNHAV 300
Db      268 QVLDSVNITLGADELKHAVFVRPVSVAFOVKDFRLYKSGVYSDTCGSTMDVNEAV 327
Qy      301 LAVGVGVEDGIPYWLKNSGNTWGDNGYFKMELGNMCGVATCASYPIVA 351
Db      328 LAVGVGVEDGVPWLKNSGADWDGHDGYFKMELGNMCGVATCASYPIVA 378

RESULT 10
US-10-087-714-22
; Sequence 22, Application US/10087714
; Publication No. US20030070188A1
; GENERAL INFORMATION:
; APPLICANT: Havkin-Frenkel, Daphna
; APPLICANT: Podolski, Andrzej
; APPLICANT: Dixon, Richard A.
; TITLE OF INVENTION: Vanillin Biosynthetic Pathway Enzyme From Vanilla Planifolia
; FILE REFERENCE: DMCI0039
; CURRENT APPLICATION NUMBER: US/10/087,714
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: 09/462,576
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: PCT/US98/14895
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: 60/052,604
; PRIOR FILING DATE: 1997-07-15
; PRIOR APPLICATION NUMBER: 60/272,415
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Nicotiana tobaccum
US-10-087-714-22

Query Match      69.0%; Score 1286; DB 14; Length 360;
Best Local Similarity 69.6%; Pred No. 5, 8e-127;
Matches 247; Conservative 35; Mismatches 65; Indels 8; Gaps 5;

Qy      5 LLEFLLFLVSALSVALAG---FEEDNPIRS-VTQRPDSIEPAILGVLSGCRHAFHAFPA 60
Db      6 LLLALVAGGLFASALAGPATADENPIRVQVSDGLHELENAILQVVGKTRHLSAFRFA 65

Qy      61 RYKSGSYGSEBEIKKRFGI FVENLAFTSTNRKDLSTYLTGINOPADLTWEEFRTNRLCA 120
Db      66 HRYGKRYSEVEIKRFEVFLDNLKMIRSHNKKGLSYKLVGNFEFDTLWEEFRTNRLCA 125

Qy      121 QNCSTAHGNHREVDGVLPTVRDREQGIIVSPVKDQSCGS-WTFSTTGALAAAYTQLTG 179
Db      126 QNCSTATTGNLKVNTVNLVLPETKDWREAGIVSPVKQKCGSCWTFTSTTGALAAAYSQAF 185

Qy      180 S--TLSEQQLVDCASAFNNFGC--GLPSQAFVYKNGGIDTEQTPYLVGMGICNFKQE 236
Db      186 KGISLSEQQLVDCAGAFNFGCGGLPSQAFVYKNGGLDTEEAIPYTKGNGLCKFSE 245

Qy      237 NVGVKVIDSINITLGADELKHAVGLVRPVSVAFEVVKGNLYKKGVSDDTCGRDPMV 296
Db      246 NVGVKVIDSVNITLGADELKHAVGLVRPVSVAFEVVKGNLYKKGVSDDTCGRDPMV 305

Qy      297 NHAVLAVGVGVEDGIPYWLKNSGNTWGDNGYFKMELGNMCGVATCASYPIVA 351

```

```

Db      306 NHAVLAVGVGVEDGVPWLKNSGADWDGDEGIFKMEKGNMCGVATCASYPIVA 360

RESULT 11
US-10-424-599-195993
; Sequence 195993, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 195993
; LENGTH: 353
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_19007C.1.pap
US-10-424-599-195993

Query Match      68.6%; Score 1280; DB 12; Length 353;
Best Local Similarity 68.3%; Pred No. 2, 4e-126;
Matches 244; Conservative 34; Mismatches 69; Indels 10; Gaps 5;

Qy      1 MAAKLLFFLLFLVSALSVALAG--FEEDNPIRSVTQRPDSIEPAILGVLSGCRHAFHAF 58
Db      1 MARLSLLIFAFCAVAVAVAGSSFFDANPRLAS---DLESQVLVDVIGQSRHLSAFAR 56

Qy      59 FARRYKSGSYGSEBEIKKRFGI FVENLAFTSTNRKDLSTYLTGINOPADLTWEEFRTNRLG 118
Db      57 FARRHGKRYSEVEIKRFEVFLDNLKMIRSHNKKGLSYKLVGNFEFDTLWEEFRTNRLG 116

Qy      119 AAQNCSTAHGNHREVDGVLPTVRDREQGIIVSPVKDQSCGS-WTFSTTGALAAAYTQ 177
Db      117 APQNCSTATLKGHRLTDAVLPEKDKWRKEGIVSQVQKCGSCWTFTSTTGALAAAYQA 176

Qy      178 TGS--TLSEQQLVDCASAFNNFGC--GLPSQAFVYKNGGIDTEQTPYLVGMGICNFK 234
Db      177 FGKNI LSEQQLVDCAGAFNFGCGGLPSQAFVYKNGGLDTEEAIPYTKGNGLCKFT 236

Qy      235 QENVGVKVIDSINITLGADELKHAVGLVRPVSVAFEVVKGNLYKKGVSDDTCGRDPM 294
Db      237 AKXAVRVVIDSINITLGADELKHAVGLVRPVSVAFEVVKGNLYKKGVSDDTCGRDPM 296

Qy      295 DVNHAVLAVGVGVEDGIPYWLKNSGNTWGDNGYFKMELGNMCGVATCASYPIVA 351
Db      297 DVNHAVLAVGVGVEDGVPYWIHKNSGNSWGDNGYFKMELGNMCGVATCASYPIVA 353

RESULT 12
US-10-425-114-44426
; Sequence 44426, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128

```

```

; SEQ ID NO 44426
; LENGTH: 353
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700980869_FLI pep
US-10-425-114-44426

Query Match      68.6%; Score 1280; DB 12; Length 353;
Best Local Similarity 68.3%; Pred. No. 2.4e-126;
Matches 244; Conservative 34; Mismatches 69; Indels 10; Gaps 5;

Qy 1 MAKKLLPFLVLSALVALAG--FEEDNPIRSVTQRPDSEIPAILGVLSGCHAFHFA 58
Db 1 MARLSLLIFAPCAVAVAVAGSFFDDANPIRLAS-----DLESOVLVDVIGSRRHLSFA 56

Qy 59 FARRYGKSYGSEEBIKRFGIFVENLAFIRSTNRKDLSTYTLGINQFADLTWEEFRNRLG 118
Db 57 FARRHGKRYRSDVIRNFRIFSDNLKIRSTNERSLTYTLGVNHFPADTWEEFRHKL 116

Qy 119 AAKNCSTAHGNHRFVDGLPVTRDREOGIVSPVKDQSGCS-WTFSITGALAAAYTQ 177
Db 117 APQNCSTATLKNHRLTDVLPDDEKDRKEGIVSQVKDQSGCSWTFSITGALAAAYQA 176

Qy 178 TGS--TLSEQQLVDCASAFNNFGC--GGLPSQAFYVKYNGIGDTEQTPYVLGYMGICNFK 234
Db 177 FGKISLSEQQLVDCAGAFNNFGCGGLPSQAFYIKYNGIGLDEEAYPTGKGVCKFT 236

Qy 235 QENVGVKVIDSINITLGADELKHAVGLVRPVSVAFEVVKGNLYKKGVMCGVATCASYPIVA 294
Db 237 AKNAVAVRVIDSINITLGADELKQAVFVRPVSVAFEVAVKDFRFFYNGVYTSTICGSTM 296

Qy 295 DVNHAVLAGYGVGVEDGIPYWLIIKNSWGTNMGDNGYFKMELGKNMCGVATCASYPIVA 351
Db 297 DVNHAVLAGYGVGVEDGVPIYWLIIKNSWGTNMGDNGYFKMELGKNMCGVATCASYPIVA 353

RESULT 13
US-10-087-714-20
; Sequence 20, Application US/10087714
; Publication No. US20030070188A1
; GENERAL INFORMATION:
; APPLICANT: Havkin-Frenkel, Daphna
; APPLICANT: Podstolski, Andrzej
; TITLE OF INVENTION: Vanillin Biosynthetic Pathway Enzyme From Vanilla Planifolia
; FILE REFERENCE: DMCI0099
; CURRENT APPLICATION NUMBER: US/10/087,714
; PRIOR FILING DATE: 2002-02-28
; PRIOR FILING DATE: 1997-07-15
; PRIOR FILING DATE: 2000-05-22
; PRIOR FILING DATE: 1998-07-15
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-087-714-20

Query Match      68.2%; Score 1272.5; DB 14; Length 360;
Best Local Similarity 66.4%; Pred. No. 1.6e-125;
Matches 239; Conservative 40; Mismatches 72; Indels 9; Gaps 4;

Qy 1 MAKKLLF---FLLFLVLSALVALAGFEEDNPIRSVTQRPDSEIPAILGVLSGCHAFHFA 57
Db 1 MSAKTLSSWLVLVLAASAAANIGFDESNPITRMVSDGLREVESVSQILGQSRHVLFA 60

; SEQ ID NO 44426
; LENGTH: 353
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700980869_FLI pep
US-10-425-114-44426

Query Match      68.6%; Score 1280; DB 12; Length 353;
Best Local Similarity 68.3%; Pred. No. 2.4e-126;
Matches 244; Conservative 34; Mismatches 69; Indels 10; Gaps 5;

Qy 1 MAKKLLPFLVLSALVALAG--FEEDNPIRSVTQRPDSEIPAILGVLSGCHAFHFA 58
Db 1 MARLSLLIFAPCAVAVAVAGSFFDDANPIRLAS-----DLESOVLVDVIGSRRHLSFA 56

Qy 59 FARRYGKSYGSEEBIKRFGIFVENLAFIRSTNRKDLSTYTLGINQFADLTWEEFRNRLG 118
Db 57 FARRHGKRYRSDVIRNFRIFSDNLKIRSTNERSLTYTLGVNHFPADTWEEFRHKL 116

Qy 119 AAKNCSTAHGNHRFVDGLPVTRDREOGIVSPVKDQSGCS-WTFSITGALAAAYTQ 177
Db 117 APQNCSTATLKNHRLTDVLPDDEKDRKEGIVSQVKDQSGCSWTFSITGALAAAYQA 176

Qy 178 TGS--TLSEQQLVDCASAFNNFGC--GGLPSQAFYVKYNGIGDTEQTPYVLGYMGICNFK 234
Db 177 FGKISLSEQQLVDCAGAFNNFGCGGLPSQAFYIKYNGIGLDEEAYPTGKGVCKFT 236

Qy 235 QENVGVKVIDSINITLGADELKHAVGLVRPVSVAFEVVKGNLYKKGVMCGVATCASYPIVA 294
Db 237 AKNAVAVRVIDSINITLGADELKQAVFVRPVSVAFEVAVKDFRFFYNGVYTSTICGSTM 296

Qy 295 DVNHAVLAGYGVGVEDGIPYWLIIKNSWGTNMGDNGYFKMELGKNMCGVATCASYPIVA 351
Db 297 DVNHAVLAGYGVGVEDGVPIYWLIIKNSWGTNMGDNGYFKMELGKNMCGVATCASYPIVA 353

RESULT 13
US-10-087-714-20
; Sequence 20, Application US/10087714
; Publication No. US20030070188A1
; GENERAL INFORMATION:
; APPLICANT: Havkin-Frenkel, Daphna
; APPLICANT: Podstolski, Andrzej
; TITLE OF INVENTION: Vanillin Biosynthetic Pathway Enzyme From Vanilla Planifolia
; FILE REFERENCE: DMCI0099
; CURRENT APPLICATION NUMBER: US/10/087,714
; PRIOR FILING DATE: 2002-02-28
; PRIOR FILING DATE: 1997-07-15
; PRIOR FILING DATE: 2000-05-22
; PRIOR FILING DATE: 1998-07-15
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-087-714-20

Query Match      67.6%; Score 1261.5; DB 14; Length 362;
Best Local Similarity 66.3%; Pred. No. 2.3e-124;
Matches 240; Conservative 44; Mismatches 67; Indels 11; Gaps 6;

Qy 1 MAKKLLFLLFL-----VSALSVALAGFEEDNPIRSVTQRPDS--IEPAILGVLSGCHAF 54
Db 1 MAHRILLJLJLAAVAAVATSAVAAASGPDSPISRTVTHAASALESTVIAALGRTRGAL 60

Qy 55 HFARFARRYKSYGSEEBIKRFGIFVENLAFIRSTNRKDLSTYTLGINQFADLTWEEFR 114
Db 61 RFARFARRYKRYGDAAEVQRRFRIFPSESLELVRSTNRRLGYPYRLGINRPFADMSWEFQA 120

Qy 115 NRIAGAACNCSATAHGNHRFVDG--VLPVTRDREOGIVSPVKDQSGCS-WTFSITGALAA 172
Db 121 SRLGAAQNCSTATLGNHRMRDAPALPETKDWEDGIVSPVKDQGHGCSWPFSTTGSLEA 180

Qy 173 AYTQLTGS--TUSEQQLVDCASAFNNFGC--GGLPSQAFYVKYNGIGDTEQTPYVLGYMG 229
Db 181 RYVQATGPPVLSLSEQQLADCATRYNNFSGSGGLPSQAFYIYKNGGLDTEAYPYTVGNG 240

Qy 230 ICNFKQENVGVKVIDSINITLGADELKHAVGLVRPVSVAFEVVKGNLYKKGVSSTDC 289
Db 241 ICHYPENAGHKVLDVSNITLVADELKNAVGLVRPVSVAFQVINGFRMYKSGVYTSDDC 300

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 29, 2004, 22:17:04 ; Search time 3294 Seconds
(without alignments)
4631.678 Million cell updates/sec

Title: US-10-087-714-2

Perfect score: 1865
Sequence: 1 MAAKLLFLLPLVLSALVAL.....ELGKMGCVATCASYPIVAV 352

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q/cgn2_1/USFTO_spool/US10087714/runat_26032004_113452_933/app.query.fasta_1.519
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HSPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10087714@cgn 1 1 2372 arunat_26032004_113452_933 -NCPUS=6 -ICPU=3
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.man.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1322.5	70.9	1382	8	MZECYPA D45403 Corn mRNA f
2	1320.5	70.8	1486	8	AK066748 Oryza sat
3	1310.5	70.3	1412	8	ZMSEE1 X99936 Z.mays mRNA
4	1310.5	70.3	1442	6	A43549 Sequence 17
5	1300.5	69.7	1407	6	AR231148 Sequence
6	1297	69.5	1332	8	AF454960 Brassica
7	1286	69.0	1340	8	AB032168 Nicotiana
8	1282	68.7	1345	8	LMU249847 Lolium mu
9	1281.5	68.7	1468	8	RICOZC D90408 Oryza sativ
10	1279.5	68.6	1305	8	AF233883 Arabidops
11	1279.5	68.6	1331	8	BT000676 Arabidops
12	1279.5	68.6	1352	8	AY088662 Arabidops
13	1279.5	68.6	1355	8	AF083703 Arabidops
14	1279.5	68.6	1370	8	AF360273 Arabidops
15	1279.5	68.3	1409	8	BT000673 Arabidops
16	1274.5	68.3	1366	8	BT000674 Arabidops
17	1265.5	67.9	1402	8	AY091771 Arabidops
18	1262.5	67.7	1347	8	PHU31094 Petunia hyb
19	1261	67.6	1444	8	PAU33166 Prunus arme
20	1254.5	67.3	1381	8	PSRNACP Z68291 P.sativum m
21	1252.5	67.2	1361	8	PSA278699 Pisum sat
22	1249	67.0	1388	6	A43538 Sequence 6
23	1249	67.0	1388	8	LECYP3GN 248736 L.esculentu
24	1204	64.6	1484	8	CACG2 X93220 Cicer ariet
25	1000	53.6	91274	8	ATF18N11 AL132953 Arabidops
26	929	49.8	77298	8	AB011483 Arabidops
27	822.5	44.1	2966	8	AY017412 Arabidops
28	796	42.7	1362	10	RNCATH AY017412 Nicotiana
29	792	42.5	1149	4	SSAF001169 Y00708 Rat mRNA fo
30	783.5	42.0	937	8	AY383240 Glycine m
31	779	41.8	1360	10	RATCATH M38135 Rat catheps
32	765	41.0	1381	10	MCCTSH U06119 Mus musculu
33	763	40.9	972	9	AF426248 Homo sapi
34	763	40.9	1008	9	AF426247 Homo sapi
35	763	40.9	1372	10	BC006878 Mus muscu
36	763	40.9	1399	9	HSCATH X16832 Human mRNA
37	759	40.7	1470	9	BC002479 Homo sapi
38	757	40.6	1438	9	AK130158 Homo sapi
39	744	39.9	14780	8	AP005419 Oryza sat
40	722	38.7	536	8	AF213939 Prunus du
41	695	37.3	5372	8	HVLEU X05167 Barley Gene
42	691	37.1	1106	6	AX774741 Sequence
43	691	37.1	1106	9	HSCATH X07549 Human mRNA
44	671	36.0	1410	8	AK109136 Oryza sat
45	661.5	35.5	1168	3	AY336797 Rhipiceph

ALIGNMENTS

RESULT 1

MZECYPA 1382 bp mRNA linear PLN 10-FEB-1999
 LOCUS
 DEFINITION Corn mRNA for cysteine proteinase, clone CCP2, complete cds.
 ACCESSION D45403
 VERSION D45403.1 GI:1688044
 KEYWORDS cysteine proteinase.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
 REFERENCE 1 (bases 1 to 1382)
 Domoto, C., Watanabe, H., Abe, M., Abe, K. and Arai, S.
 Isolation and characterization of two distinct cDNA clones encoding corn seed cysteine proteinases
 JOURNAL Biochim. Biophys. Acta 1263 (3), 241-244 (1995)
 MEDLINE 96004895
 PUBMED 7548211
 REFERENCE 2 (bases 1 to 1382)
 Domoto, C.
 Direct Submission
 TITLE Submitted (31-JAN-1995) Chieko Domoto, Atomi Junior College; Ohtsuka 1-5-2, Bunkyo-ku, Tokyo 112, Japan
 JOURNAL (Tel: 03-3941-8161 (ex. 575), Fax: 03-3945-1836)
 COMMENT On Nov 27, 1996 this sequence version replaced gi:644489.
 Sequence updated (26-Oct-1996) by: Chieko Domoto.

FEATURES

Location/Qualifiers
 1..1382
 /organism="Zea mays"
 /mol_type="mRNA"
 /db_xref="taxon:4577"
 /clone="CCP2"
 52..1134
 /codon_start=1
 /product="cysteine proteinase"
 /protein_id="BAA08245.1"
 /db_xref="GI:1644490"
 /translation="MVPRLFLVAVLVADTAAVNSGFADSNPIRPVTDRAALES
 TVFAALGRDALARFARFVRVYKSVESAEVHVRIFSESLQIVRSNKRGLSVRL
 GINRFADMSHEPRALGNAONCSATLGNHMRMAAVALPETKDWREDGIVSPVKN
 QHGCSTWSTTGALAEATQATKPISELOLVDCGFANFNCGNGLSPQAFY
 IKYNGGLDTEESTYPCGVNGICFKFNENGVGRVDSVNITLGADELKDAVLGRPV
 VAPEVITGFRLYKSGVYTSDDCTPMDVNHAVLVGVGVEDGVFWLLKNSWGADWG
 DEGYFKMEMKMKMGVATCASYPIVA"

ORIGIN

Alignment Scores:
 Pred. NO.: 7,49e-111 Length: 1382
 Score: 1322.50 Matches: 258
 Percent Similarity: 79.44% Conservative: 28
 Best Local Similarity: 71.67% Mismatches: 65
 Query Match: 70.91% Indels: 9
 DB: 8 Gaps: 6

US-10-087-714-2 (1-352) x MZECYPA (1-1382)

Qy 1 MetAlaAlaLysLeuLeuPheLeuPheLeuValSerAlaLeuSerValAlaLeu 20
 Db 52 ATGTCCTCCACCGCGCTGTCCTCGCGCTGTCCTCGCGGACACCGCGCGCTG 111
 Qy 21 -----AlaGluPheGluGluAspAspProIleArgSerValThrGlnArgProAspSer 38
 Db 112 GTCAACTCCCGCTTCGGCGCTCCACACCGATCCCGCGCTCCCGCGCGCGCGCTCC 171
 Qy 39 ---IleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheAla 57
 Db 172 CGCTCGAGTCCACGGTCTTCGCGCGCTTCGCGCGCGCGCGCGCGCGCTTCGCA 231
 Qy 58 ArgPheAlaArgArgTyrGlyValSerTyrGlySerGluGluIleLysLysArgPhe 77
 Db 232 CGCTTCGCGCGTACGAGTACGCAAGCTACGAGCGCGCGCGCGCGCGCGCTTC 291
 Qy 78 GlyIlePheValGluGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyr 97

Db 292 AGGATCTTCTCCGAGAGCTCCAGCTGTCCTCGCTCCACCAACCGCAAGCGCTCTCTCTAC 351
 Qy 98 ThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeu 117
 Db 352 CGCTTCGCGATCAACCGCTTCGCGGACATGATGCTGGGAGGAGTTCGTCGCGCGCTC 411
 Qy 118 GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPhe-----ValAsp 135
 Db 412 GCGCAGCCAGCACTGCTCCGCGACCTTACCGCGCAACACCGGATGCGCGCGCGCGCC 471
 Qy 136 GlyValLeuProValThrArgAspTrpArgGluGlnGlyIleValSerProValLysAsp 155
 Db 472 GTTCGCGTCCGAGAGCAAGAGCTGCGAGGAGGATGGGATTTGAGCCAGTGAAAC 531
 Qy 156 GlnGlySerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaIleTyr 174
 Db 532 CAGGCGCACTGTGATCATGTGGACCTTCCAGCACTACTGGTGCATTTGAGGCTGCATAT 591
 Qy 175 ThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSer 192
 Db 592 ACCAGGCACTGCGAAGCCCATCTCTCTCTGAGCAACAGCTTGTGACTGTGGTTTT 651
 Qy 193 AlaPheAsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLys 211
 Db 652 GCATTCAACAATTTCCGATGCAACGAGGCGCTTCCATCCAGGCGCTTTGATATACATCAA 711
 Qy 212 TyrAsnGlyGlyIleAsnThrGluGlnThrTyrProTyrLeuGlyValMetGlyIleCys 231
 Db 712 TACATGTGGCTTGACATGAGGATCTTACCTTACCAAGGTGTCAATGGAATCTGC 771
 Qy 232 AsnPhelysGlnGluAsnValGlyValLysValIleAspSerIleAsnIleThrLeuGly 251
 Db 772 AAGTTTAAGAAATGAGATGTTGGAGTCAAGGTTTGGACTCGTAAACATCACCTCGGT 831
 Qy 252 AlaGluAspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGlu 271
 Db 832 GCTGAGGATGACCTGAGGATGCTGTGGTCTGGTTCGCCAGTAGTGTTCCTTCGAG 891
 Qy 272 ValValLysGlyPheAsnLeuTyrIleLysGlyValTyrSerSerAspThrCysGlyArg 291
 Db 892 GTGATCACTGGTTTCAGGCTGTACAGAGCGGAGTTTACACTAGCGACCATTTGGAACT 951
 Qy 292 AspProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIle 311
 Db 952 ACACCGATGGATGTGAACCGCTGTTCTGGCTGTGGCTACGGTGTGGAAGTGGTGA 1011
 Qy 312 ProTyrTrpLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLys 331
 Db 1012 CCTACTGGCTCATCAAGAACTCATGGGCGCTGACTGGGCTGATGAGGGTTACTTCAAG 1071
 Qy 332 MetGluLeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIleValAla 351
 Db 1072 ATGGAATGGGCAAGACATGTGCGGTGTTGCTAGTGTGCTATCTACCTACCTATTTCGCA 1131

RESULT 2

AK066748 1486 bp mRNA linear PLN 24-JUL-2003
 LOCUS Oryza sativa (japonica cultivar-group) cDNA clone J013074D19, full insert sequence.
 DEFINITION
 ACCESSION AK066748
 VERSION AK066748.1 GI:32976766
 KEYWORDS FLI CDNA; CAP trapper.
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzae; Oryza.

REFERENCE

1
 The Rice Full-Length cDNA Consortium, National Institute of Agricultural Sciences Rice Full-Length cDNA Project Team., Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,

Db 68 CTGCTGATTCATCGCCGCGATCGACGGCGGAGAGATATCGGATTGTGATGACTCAACCCG 127
 Qy 29 lleAgservValThrGluArgProAspSerleGluProAlaIleLeuGlyValleuGly 48
 Db 128 ATCCGTATGGTCTCCGACGGTCTCCGGAGGTAGAGAAATCGGTGTCCAGATCTTAGC 187
 Qy 49 SerCysArgHisAlaPheHisPheAlaArgPheAlaArgAgtTyrGlyLysSerTyrGly 68
 Db 198 CAATCAGCCAGGTATCTCTTCGCTCGCTCGCTCAGTACAGTACGGAGAGATCAGG 247
 Qy 59 SerGluGluGluIleLysLysArgPheGlylePheValGluAsnLeuAlaPheIleArg 88
 Db 248 AACCGGAGGAGATCAAGCTCAGGTTCGATATTCAGAGAGATCTGGATTTCATCAGA 307
 Qy 89 SerThrAsnArgLysAspLeuSerTyrThrLeuGlyIleAsnGlnPheAlaAspLeuThr 108
 Db 308 TCCACCAATAGAAGGCTTGTCTTACAACTCGGTGTCAATCAGTTTGTGTATATGACG 367
 Qy 109 TrpGluGluPheArgThrAsnArgLeuGlyAlaAlaGlnAsnCysSerAlaThrAlaHis 128
 Db 368 TGGCAAGAGTTTCAAGGACCAAGCTTGTGTCTCAAACTGCTCTGCCACTTTAAAG 427
 Qy 129 GlyAsnHisArgPheValAspGlyValLeuProValThrArgAspTTPArgGluGly 148
 Db 428 GAAACCCCAAACTCACAGGAGAGCTCTTCGGAAACCAAGACTGGAGAGAGATGGT 487
 Qy 149 IleValSerProValLysAspGlnGlySerCysGlySer---TrpThrPheSerThrThr 167
 Db 488 ATTGTAGCCCTGTCAAAGACAGGGGGTGTGTGATCGTCTGACATTCAGTACTACT 547
 Qy 168 GlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGlySer-----ThrLeuSerGluGln 185
 Db 548 GGAGCGCTTGGAGCGAGCTTATCACCAAGCATTTGGAAAGGAATATCACTCTCTGAGCAA 607
 Qy 186 GlnLeuValAspCysAlaSerAlaPheAsnAsnPheGlyCys---GlyGlyLeuProSer 204
 Db 608 CAGCTTGTGATTTGTCTGGAGCTTCAACAACTATGCTGCAATGGTGGCTTCTCTCC 667
 Qy 205 GlnAlaPheGluTyrValLysTyrAsnGlyGlyIleAspThrGluGlnThrTyrProTyr 224
 Db 668 CAAGCTTTGAATACATCAAACTCCAAACGGCGGCTCGACAGCGAGGAGCTTATCTCTAC 727
 Qy 225 LeuGlyValMetGlyIleCysAsnPheLysGlnGluAsnValGlyValLysValIleAsp 244
 Db 728 ACCGTTGAGATGGAACTGCAAGTATCTCTGAAACGTCGGTGTAGAGTCTCTGAC 787
 Qy 245 SerIleAsnIleThrLeuGlyAlaGluAspGluLeuLysHisAlaValGlyLeuValArg 264
 Db 788 TCAGTCAACATTAATCTCGGTGCTGAAGATGAATGAAGCATGCGGTGGATTGGTACGA 847
 Qy 265 ProValSerValAlaPheGluValLysGlyPheAsnLeuTyrLysGlyValTyr 284
 Db 848 CAGTACAGATGATTCAGATTATACCTCGCTTCGGCTTTTACAGAGCGGAGTTTAC 907
 Qy 285 SerSerAspThrCysGlyArgAspProMetAspValAsnHisAlaValLeuAlaValGly 304
 Db 908 AGCGATAGTCACTGTGGACAGACTCCAATGACGCTGAACCACTGTTTGGCGGTGGT 967
 Qy 305 TyrGlyValGluAspGlyIleProTyrTrpLeuIleLysAsnSerTrpGlyThrAsnTrp 324
 Db 968 TATGAATTGAGACGGTGTCCATATTGGCTTTATAAGAACTCATGGGAGCAGATGG 1027
 Qy 325 GlyAspAsnGlyTyrPheLysMetGluLeuGlyLysAsnMetCysGlyValAlaThrCys 344
 Db 1028 GGTGACAAAGTTACTTCAAGATGGAGATGGGAGAACATGTGTGTATGGACATGT 1087
 Qy 345 AlaSerTyrProIleValAla 351
 Db 1088 GCATCGTACCGGTTGTGGCC 1108
 RESULT 7
 AB032168
 LOCUS AB032168 1340 bp mRNA linear PLN 24-JAN-2001

DEFINITION Nicotiana tabacum NTCP-23 mRNA for cysteine protease, complete cds.
 ACCESSION AB032168
 VERSION AB032168.1 GI:8347419
 KEYWORDS cysteine protease.
 SOURCE Nicotiana tabacum (common tobacco)
 ORGANISM Nicotiana tabacum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamids; Solanales; Solanaceae; Nicotiana.
 REFERENCE 1 (sites)
 AUTHORS Ueda,T., Seo,S., Ohashi,Y. and Hashimoto,J.
 TITLE Circadian and senescence-enhanced expression of a tobacco cysteine protease gene
 JOURNAL Plant Mol. Biol. 44 (5), 649-657 (2000)
 MEDLINE 21039173
 PUBMED 11198425
 REFERENCE 2 (bases 1 to 1340)
 AUTHORS Ueda,T.
 TITLE Direct Submission
 JOURNAL Submitted (03-SEP-1999) Tadamasu Ueda, National Institute of Agrobiological Resources, Department of Molecular Genetics, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail:uechueabr@affrc.go.jp, Tel:81-297-38-7443, Fax:81-298-38-7044)
 FEATURES
 source
 1..1340
 /organism="Nicotiana tabacum"
 /mol_type="mRNA"
 /cultivar="Sameun NN"
 /db_xref="taxon:4097"
 /clone="NTCP-23"
 /tissue_type="leaf"
 /clone_lib="Lambda ZAP healthy tobacco"
 /tissue_lib="healthy leaf"
 gene
 1..1340
 /gene="NTCP-23"
 CDS
 35..1117
 /gene="NTCP-23"
 /codon_start=1
 /product="cysteine protease"
 /protein_id="BAA9501.1"
 /db_xref="GI:8347420"
 /translation="MSRPSLLALVVAGLPASALAGPATADENPIROVSDGLHEL
 ENALIQVQKTRHALSPARFAHYKRYESVEIKQREVFLDNLKMRSHNKGILSY
 KIGNEFTLTWDFRRDLRGAQNCSTATKGNLKVTVLVPETKDRRAGLIVSPYKN
 QKQSCSWTFTGALRAYSQAFKGLSLSQQLVDCAGAFNFCNGGLPSQAFY
 IKSNGLDTEAYPTGKNGLCFKFSENVGVKVIDSVNITLGADELKVALVRPVS
 IAFYIKGPKQYKGVYSTECCNTPMDVNHAVLAVGVGVENGVPYWLKNSWGDWG
 DNGYFKWGMGKNCIGIATCATSPVVA"
 125..142
 /gene="NTCP-23"
 /note="vacuolar sorting sequence"
 461..1114
 /gene="NTCP-23"
 /product="cysteine protease"
 /function="protein degradation"
 /note="putative"
 sig_peptide
 125..142
 /gene="NTCP-23"
 mat_peptide
 461..1114
 /gene="NTCP-23"
 /note="putative"
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,51e-107 Length: 1340
 Score: 1286.00 Matches: 246
 Percent Similarity: 79.72% Conservative: 37
 Best Local Similarity: 69.30% Mismatches: 64
 Query Match: 68.95% Indels: 8
 DB: Gaps: 5
 US-10-087-714-2 (1-352) x AB032168 (1-1340)
 Qy 5 LeuLeuPhePheLeuPheLeuValSerAlaLeuSerValAlaLeuAlaGly----- 22
 Db 50 CTCCTATTGGCTCTGCTGCTCGCGGTGGCCCTTTTCGCTCGCACTCGCCGACCGCGG 109

Qy 23 ---PheGluGluAspAsnProIleArgSer---ValThrGlnArgProAspSerIleGlu 40
 Db 110 ACCTTTGCCGATGAGATCCGATCAGACAAAGTCGTTCTCAGCGTTTACATGAGTCGAG 169
 Qy 41 ProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAla 60
 Db 170 AAGCAATCTCCAAAGTCGTCGCAAGACCCGCGATGCTCTCTCTCTCTCTCTCTCTCT 229
 Qy 61 ArgArgTyrGlyLysSerTyrGlySerGluGluGluIleLysLysArgPheGlyIlePhe 80
 Db 230 CACAGGTATGGAGAGTACGATCAGTTCAGTTCAGGAGATAAAGCAAGGTCGAGGTATT 289
 Qy 81 ValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyrThrLeuGly 100
 Db 290 TTGACAAATTGAGATGATTCGATCGCACAACAAAGAGAGTATCATACAAACTCGGT 349
 Qy 101 IleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAlaAla 120
 Db 350 GTCATGAGTTACCGACCTTAACATGGACGAGTTCGGAGACAGGTTGGGGCAGCT 409
 Qy 121 GlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProVal 140
 Db 410 CAAACATGTTCCAGCCACCAAAAGGCAATCTCAAAGTCACTAACGTTGTTCTGCCGGAG 469
 Qy 141 ThrArgAspTrpArgGluGlnGlyIleValSerProValLysAspGlnGlySerCysGly 160
 Db 470 ACGAAGACATGGCGGAGCTGGGATGTCAGCCAGTCAAGAACACGAGCAGTCGGA 529
 Qy 161 Ser---TrpThrPheSerThrThrGlyAlaLeuGluAlaIleThrGlnLeuThrGly 179
 Db 530 TCTTGTGGCATTCAGACATTCAGTGTGTCACATGAGTGTGAGTGTGAGTGTGAGTGT 589
 Qy 180 Ser-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPhe 197
 Db 590 AAGGAATCTCTTATCTGACGACGACTTGTGACCTGTGAGCTGTGAGCTTTTAATACTTT 649
 Qy 198 GlyCys---GlyGlyLeuProSerGlnAlaPheGluThrValLysTyrAsnGlyGlyIle 216
 Db 650 GGCTGCAATGGTGGCTCCCATCAACAGCTTGTGAGTATATAATCCATGCTGCTCT 709
 Qy 217 AspThrGluGlnThrTrpTyrProTyrLeuGlyValMetGlyIleCysAsnPheLysGlnGlu 236
 Db 710 GACACTGAAGAGCATATCCATACACCTGGCAAGATGGCTATGTAATTTCTCATCAGAA 769
 Qy 237 AsnValGlyValLysValIleAspSerIleAsnIleThrLeuGlyAlaGluAspGluLeu 256
 Db 770 AATGTTGGTGTCAAGTCATCGATTCGTCATATTACCTGGGTGCTGAAGATGAAC 829
 Qy 257 LysHisAlaValGlyLeuValArgProValSerValAlaPheGluValLysGlyPhe 276
 Db 830 AAATACCGGTTGCAATGGTTAGCCCGCTTAGTATAGCTTTTGGAGTGATAAAGGTTTC 889
 Qy 277 AsnLeuTyrLysLysGlyValTyrSerSerAspThrCysGlyArgAspProMetAspVal 296
 Db 890 AAACAATACAGAGTGGTGTTCACAGACCCGAATGGCGCAACATCCCATGATGTA 949
 Qy 297 AsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIleProTyrTrpLeuIle 316
 Db 950 AACCATGTGTTCTTCTGCTGGGTACGGTGTGAAAATGGTGTGCTTATGCTCATC 1009
 Qy 317 LysAsnSerTrpGlyThrAsnTrpGlyLysAsnGlyTyrPheLysMetGluLeuGlyLys 336
 Db 1010 AAGAAATTCATGGGAGCATGGTGGTGCACATGATGATCTTCAAAATGGAGATGGGAAG 1069
 Qy 337 AsnMetCysGlyValAlaThrCysAlaSerTyrProIleValAla 351
 Db 1070 AACATGTGTGATTGCGCATCTGCGATCTTACCCTGCTGCTGCTGCC 1114

RESULT 8

LMU249847 1345 bp mRNA linear PLN 30-JUN-2000
 LOCUS
 DEFINITION Lolium multiflorum mRNA for cysteine protease (see gene).
 ACCESSION AJ249847

VERSION AJ249847.1 GI:6851029
 KEYWORDS cysteine protease; see1 gene.
 SOURCE Lolium multiflorum (Italian ryegrass)
 ORGANISM Lolium multiflorum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Pooidae; Poae; Lolium.
 REFERENCE 1
 AUTHORS Li,Q., Bettany,A.J., Donnison,I., Griffiths,C.M., Thomas,H. and Scott,I.M.
 TITLE Characterisation of a cysteine protease cDNA from Lolium multiflorum leaves and its expression during senescence and cytokinin treatment
 JOURNAL Biochim. Biophys. Acta 1492 (1), 233-236 (2000)
 MEDLINE 20461783
 PUBMED 11004495
 REFERENCE 2 (bases 1 to 1345)
 AUTHORS Li,Q.
 TITLE Direct Submission
 JOURNAL Submitted (30-SEP-1999) Li Q., Department of cell Biology, Institute of Grassland and Environmental Research, Plas Gogerddan, Aberystwyth, Ceredigion, SY23 3EB, UNITED KINGDOM
 FEATURES
 source
 1..1345
 /organism="Lolium multiflorum"
 /mol_type="mRNA"
 /db_xref="taxon:4521"
 1..1345
 /gene="see1"
 36..1115
 /gene="see1"
 /function="essential for proteolysis"
 /codon_start=1
 /product="cysteine protease"
 /protein_id="CAB71032.1"
 /db_xref="GI:6851030"
 /db_xref="Gene:Q9M4E5"
 /db_xref="SPTREMBL:Q9M4E5"
 /translation="MAEGRILVLALAVLATAVAASPADSNPIRPVTERAASVESTVLGLGRTHALRFARFAVRKGYSGAAEVORRFIFSESLDEVRSNRLGSLYKLGINRPDWEERFOATKGAOTCSATLGNHMDANALPETKDWREGTGVSPVKDQA SCGSDWTFSTGALEAAVTQATKNISLSEQLVDCAGAYNNPGNGGLPSQAPFIK YNGIDTESYPYKGVNGVCYKRPENRAVOVADSNITLNAEDELKNAVLVRPVSA FEVIDGFKQYKSGVTSDHCCTTDFDDNVHVLAVGYGVENGVPYWLKNSGADWGED GYFKEMGKNMCATVCASYPIIAA"

ORIGIN

Alignment Scores:
 Pred. No.: 3 51e-107 Length: 1345
 Score: 1282.00 Matches: 243
 Percent Similarity: 80.51% Conservative: 42
 Best Local Similarity: 68.64% Mismatches: 63
 Query Match: 68.74% Indels: 6
 DB: Gaps: 5
 US-10-087-714-2 (1-352) x LMU249847 (1-1345)
 Qy 4 LysLeuLeuPhePheLeuLeuPheLeuValSerAlaLeuSerValAlaLeuAlaGlyPhe 23
 Db 48 CGCATCTCTGCTTCTGCGCTCGCCCTCTGCGCACGCGCGCGCTCCCTCTT 107
 Qy 24 GluGluAspAsnProIleArgSerValThrGlnArgProAspSer---IleGluProAla 42
 Db 108 GCGACTCCAAACCCGATTCGGCCCTCACCAGCGCGCGCTCGCGGTGAGTCAACG 167
 Qy 43 IleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAlaArg 62
 Db 168 GTCTCGGCGCTCGGCGCGCACCGCCCTCCGCTTCGCGCTTCGCGCTCAGG 227
 Qy 63 TyrGlyLysSerTyrGlySerGluGluIleLysLysArgPheGlyIlePheValGlu 82
 Db 228 CACGGCAAGAGTACGGGAGCGCGGAGGTCGAAAGCGGTTCCGATCTTCTCCGAG 287

Qy 115 AsnArgLeuGlyAlaAlaGlnAsnCySerAlaThraHisGlyAenHisArgPheVal 134
 Db : : : : :
 Qy 407 AGCCGGTTCGGCGGGCGAGAACTGCTCGGACGCTCCCGGCAACACCGGATGCCG 466
 Db : : : : :
 Qy 135 AspGly---ValLeuProValThrArgAspTrpArgGluGlnGlyLeValSerProVal 153
 Db : : : : :
 Qy 467 GACGCGCCGCTCTCCGCGAGACCAAGACTGGAGGAGGATGGGATCGTGAGCCCGGTG 526
 Db : : : : :
 Qy 154 LysAspGlnGlySerCySgLySer---TrpThrPheSerThrThrGlyAlaLeuGluAla 172
 Db : : : : :
 Qy 527 AAAGCAGGCTCACTGGCTCTGTGGCTTGGCCCTTCAGCACCACTGGCTCTCTTGAGGCA 586
 Db : : : : :
 Qy 173 AlaTrpThrGlnLeuThrGlySer-----ThrLeuSerGluGlnLeuValAspCys 190
 Db : : : : :
 Qy 587 CGATATACTCAGGCCACTGGGCCCTGTCTCTCTTCTTCTGAGCAGCAGCTGGCTGATTGT 646
 Db : : : : :
 Qy 191 AlaserAlaPheAenAsnPhgLyCys---GlyGlyLeuProSerGlnAlaPheGluTyr 209
 Db : : : : :
 Qy 647 GCTACTCGATACAATAATTCGGATGAGTGAGGCGCTACCATCTCAGGCGCTTTGAGTAC 706
 Db : : : : :
 Qy 210 ValLysTrpAsnGlyGlyIleAspThrGluGlnThrTyrProTyrLeuGlyValMetGly 229
 Db : : : : :
 Qy 707 ATCAATACAAATGCTGGCCTTGACACTGAAGAGGCTTACCCTTACACGGGTGTCAATGGC 766
 Db : : : : :
 Qy 230 IleCysAsnPhgLySgGlnGluAsnValGlyValLysValLleAspSerIleAsnIleThr 249
 Db : : : : :
 Qy 767 ATCTGTCTATTACAAAGCCTGAAACCGTGGAGTCAAGGTTTGGACTCCGTAAACATCACC 826
 Db : : : : :
 Qy 250 LeuGlyAlaGluAspGluLeuLysHisAlaValGlyLeuValArgProValSerValAla 269
 Db : : : : :
 Qy 827 CTGTTCTGAGGATGAGCTGAAGATGCTGTGGACTTGTCTCGTCCAGTTAGTGTGCC 886
 Db : : : : :
 Qy 270 PheGluValValLysGlyPheAsnLeuTyrLysGlyValTyrSerSerAspThrCys 289
 Db : : : : :
 Qy 887 TTTGAGTGTATCAACGGTTTCAGAGTATCAAGAGTGGAGTTTACACAAAGTACCATTGT 946
 Db : : : : :
 Qy 290 GlyArgAspProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAsp 309
 Db : : : : :
 Qy 947 GGAATCTCCAAATGATGTAACACCGCTGTCTGCGCTTGGCTATGTTGCTGCAAT 1006
 Db : : : : :
 Qy 310 GlyIleProTyrTrpLeuLysLeuAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyr 329
 Db : : : : :
 Qy 1007 GCGCTTCCCTACTCGCTCATCAAGAACTCATGGGCTGCAGACTGGGTGCAATGTTTAC 1066
 Db : : : : :
 Qy 330 PheLysMetGluLeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIle 349
 Db : : : : :
 Qy 1067 TTCAGATGGAATGGGCAAGAACTGTGCGGTATTGCTACTTGGCATCTACCCATT 1126
 Db : : : : :
 Qy 350 ValAla 351
 Db : : : : :
 Db 1127 GTTGCA 1132

RESULT 10
 AF233883
 LOCUS
 DEFINITION
 Arabidopsis thaliana AALP protein mRNA, complete cds.
 ACCESSION
 AF233883
 VERSION
 AF233883.1 GI:7230639
 KEYWORDS
 Arabidopsis thaliana (thale cress)
 ORGANISM
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 1305)
 Ahmed S.U., Rojo E., Kovaleva V., Venkataraman S., Dombrowski J.E.,
 Masuoka K. and Raikhel N.V.
 The plant vacuolar sorting receptor ATELP is involved in transport
 of NH(2)-terminal propeptide-containing vacuolar proteins in
 Arabidopsis thaliana
 J. Cell Biol. 149 (7), 1335-1344 (2000)
 JOURNAL
 MEDLINE
 20330156

10871276
 2 (bases 1 to 1305)
 Raikhel N.
 Direct Submission
 Submitted (13-FEB-2000) MSU-DOE Plant Research Lab, MSU, Plant
 Biology Building, East Lansing, MI 48824, USA
 Location/Qualifiers
 1.1305
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 10.1186
 /note="similar to aleurain"
 /codon_start=1
 /product="AALP protein"
 /protein_id="AAF43041.1"
 /db_xref="GI:7230640"
 /translation="MSAKTILSSVVLVLAASAAANIGFDESPIRMVSDGLREVVE
 SVSILGOSRHVLSFARTRHYKGYKQNVEMKLFISIFKENLDLIRSTNKKGLSKYL
 GYNGFADLTWQEFORTKLAAGNCSATLKSHKYTEAALPETKDRSDGI VSPVKDQ
 CGSCWTFTTGALEAAHQAFAGISLSEQLVDCAGAFNNYSCNGGLPSQAPEYIK
 SNGGLDTEKAYPTGKDETKFSAENVGVQVINSVNTLGAEDLKAHAGLVRPVSIA
 FEVTHSRLYKSGYTTDSHCGSTPMDVNHAVLAVGYGVEDGVPYWLINSGWADWDK
 GYFKMEMKMKMGCIATCASYPVA"

ORIGIN

Alignment Scores:
 Pred. No.: 5.7e-107 Length: 1305
 Score: 1279.50 Matches: 240
 Percent Similarity: 77.93% Conservative: 39
 Best Local Similarity: 67.04% Mismatches: 72
 Query Match: 68.61% Indels: 7
 DB: 8 Gaps: 4

US-10-087-714-2 (1-352) x AF233883 (1-1305)

Qy 1 MetalAlaLysLeuLeuPhe-----PheLeuLeuPheLeuValSerAlaLeuSer 17
 Db : : : : :
 Db 10 ATGTCTCGGAAACAACTCTATCATCATGATGTTTGGTGTCTCGTCCGCGCATCAGCA 69
 Qy 18 ValAlaLeuAlaGlyPheGluAspAsnProIleArgSerValThrGlnArgProAsp 37
 Db : : : : :
 Db 70 GCGGTATATATCGGATCGATGAGTCAAAACCGATCCGATGGTCTCGATGGTCTCCGG 129
 Qy 38 SerIleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
 Db : : : : :
 Db 130 GAGTAGAAGAACTGTTTCCAGATCTTAGTCAATCTCGTCACTCTCTCTCTTCTGCT 189
 Qy 58 ArgPheAlaArgArgTyrGlyLysSerTyrGlySerGluGluGluLysLysArgPhe 77
 Db : : : : :
 Db 190 CGCTTCACTCCCGATATGTTAAAGATATCAGAACGTGGAGGAGATGAAGCTTCGATT 249
 Qy 78 GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyr 97
 Db : : : : :
 Db 250 TCGATTTTCAGGAGAACTTTGATTTGATCAGATCCACCAAGAAAGGCTTATCTTAC 309
 Qy 98 ThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeu 117
 Db : : : : :
 Db 310 AAACTCGGTGTATCAATTTGCTGATTGATGTCGCAAGAGTTTCAAGAACCCAGGCTT 369
 Qy 118 GlyAlaAlaGlnAsnCySerAlaThrAlaHisGlyAsnHisArgPheValAspGlyVal 137
 Db : : : : :
 Db 370 GTGTGCTGTGCAACTGTCTGCCACTTAAAGGCGCAGCCACAGGTTCACAGAACGCT 429
 Qy 138 LeuProValThrArgAspTrpArgGluGlnGlyLeuValSerProValLysAspGlnGly 157
 Db : : : : :
 Db 430 CTTCTCTGAAACAAAAGACTGGAGAGAGATGTTATCGTTAGTCCCGTCAAGATCAGGGA 489
 Qy 158 SerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGln 176
 Db : : : : :
 Db 490 GTTGTGGATCTTGTGTGACATTCAGCAACTGGAGCTCTTGGAGCGAGCTTACCATCAG 549
 Qy 177 LeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPhe 194

```

Db      550 GCATTTGGAAAAGGAATATCTCTCTGAGCAACAGCTTGTGGATTGCTGAGCTTTC 609
QY      195 AnAsnPhcGlyCys--GlyGlyLeuProSerGlnAlaPheGluTyVallystYrAsn 213
Db      610 ATAACTATGTTGCAATGGTGGCCCTTCTCTCAAGCCTTTGAATACATAAATCCAAAC 669
QY      214 GlyGlyLeuAsnThrGluGlnThrTyProTyLeuGlyValMetGlyLeuPheCysAsn 233
Db      670 GGTGGCTCGACACAGAGAAGCTTATCCTTATACCGGTAAAGATCAAACTGCAAAATTT 729
QY      234 LysGlnGluAsnValGlyVallyValIleAspSerIleAsnIleThrLeuGlyValGlu 253
Db      730 TCAGCTGAAGAGCTTGTGTCAAGTCCCTCACTCAGTCAACATCTCTGGTGTGCTGAA 789
QY      254 AspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGluValVal 273
Db      790 GATGAACCTGAAGCATCGGTTGGATTGTGACGCCAGTAAGCATAGCATTTGAGGTATA 849
QY      274 LysGlyPheAsnLeuTyriLysGlyValTySerSerAspThrCysGlyArgAspPro 293
Db      850 CACTCGTTCGGCTTTACAAAGAGTGGAGTTTACCTGATGATCAGTCTGGAAGTACTCCA 909
QY      294 MetAspValAsnHisAlaValLeuAlaValGlyTyriGlyValGluAspGlyIleProTy 313
Db      910 ATGATGTCAACACCGCGGTTTGGCGTTGTGTATGAGTTGAGAGCGGTACCATAT 969
QY      314 TrpLeuLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyriPheLysMetGlu 333
Db      970 TGGCTTATTAAAGACTCATGCGGAGCGGATGGCGGCGCAAAAGGTACTTCAAGATGGAG 1029
QY      334 LeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyriProIleValAla 351
Db      1030 ATGGGGAAGAAATGTTGTGTATGCTATGCTATGCTATATATATATATATATATATAT 1083

```

RESULT 11

```

BT000676
LOCUS      Arabidopsis thaliana clone RAFL08-12-G17 (R11172) putative cysteine
           proteinase AAPL (At5g60360) mRNA, complete cds.
ACCESSION BT000676.1 GI:23397073
VERSION    1
KEYWORDS   Arabidopsis thaliana (thale cress)
SOURCE     Arabidopsis thaliana
ORGANISM   Arabidopsis thaliana
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 1331)
AUTHORS   Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
           Quach,H.L., Tang,C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C.,
           Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
           Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,
           Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P.,
           Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W.,
           Ecker,J.R. and Theologis,A.

```

TITLE Arabidopsis Full Length cDNA Clones

```

JOURNAL   Arabidopsis Full Length cDNA Clones
REFERENCE 2 (bases 1 to 1331)
AUTHORS   Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
           Quach,H.L., Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C.,
           Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R.,
           Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J.,
           Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M.,
           Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K.,
           Davis,R.W., Ecker,J.R. and Theologis,A.

```

```

TITLE     Direct Submision
JOURNAL   Submitted (01-OCT-2002) Plant Gene Expression Center, 800 Buchanan
           Street, Albany, CA 94710, USA
COMMENT   RIKEN Genomic Sciences Center (GSC) members carried out the
           collection and clustering of RAFL cDNAs (Arabidopsis thaliana)
           Arabidopsis Full-length cDNA: Seki,M., Narusaka,M., Ishida,J.,
           Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,

```

Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGECC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Yu,G., Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J., Shinn,P., Palm,C.J., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Davis,R.W., Ecker,J.R. and Theologis,A.

Yamada,K. (SSP/PGECC) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGECC) contributed equally to this work as PIs.

Annotation based on January 2002 version of the Arabidopsis genome submitted to Genbank.

FEATURES

Location/Qualifiers

```

1..1331
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="RAFL08-12-G17 (R11172)"
/ecotype="Columbia"
/notes="This clone is in a modified pBluescript vector (pUC-1) as a BamHI/XhoI insert."

```

gene

```
1..1331
/gene="At5g60360"

```

5'UTR

```
1..68
/gene="At5g60360"

```

misc_difference

```
2..4
/gene="At5g60360"
/notes="compared to genomic sequence"
/replaces="gct"
69..1145
/gene="At5g60360"
/codon_start=1

```

CDS

```

69..1145
/notes="compared to genomic sequence"
/replaces="gct"
69..1145
/gene="At5g60360"
/evidence="experimental"
/product="putative cysteine proteinase AAPL"
/protein_id="AA031822.1"
/db_xref="GI:23397074"
/translation="MSAKTLLSVLVLVVAASAAANIGFDSNPIMVSDGLREVVEE
SVNQLQGRHVLVSFAFTHRYGKYQNVVEMLRFSIFKNLDLIRTNKGLSYKL
GVNQFADLTWQEFQTLGAQNCSTALGSKHKTFAALPETKMDREDGIVSPKQGG
GCSCWTFTTGALAAAYHQAQFGKISLSEQLVDGAGAFNNGCMGGLPSQAFYIK
SNGLDTERKAYPTGKDETKCFKSAENVGVQLNSVNTILGAEDLKHAGLVKRPVIA
FEVTHSFRLYKSGVTTDSHCGSTPMDVNEHVLAVGVGVEDGVYIWLKNSWGDGDK
GYFKMGMKMGICATCASPVA"
1146..1331
/gene="At5g60360"

```

3'UTR

```
1146..1331
/gene="At5g60360"

```

ORIGIN

```

Alignment Scores:
Pred. No.:      5,85e-107      Length:      1331
Score:          1279.50        Matches:     240
Percent Similarity: 77.93%      Conservative: 39
Best Local Similarity: 67.04%    Mismatches:  72
Query Match:     68.61%        Indels:      7
DB:              8             Gaps:         4

```

US-10-087-714-2 (1-352) x BT000676 (1-1331)

```

Qy      1 MetAlaAlaLysLeuLeuPhe-----PheLeuLeuPheLeuValSerAlaLeuSer 17
Db      69 ATGCTGCGAAAACAATCTCTATCATCATGATGTTTGGTGTCTCGCCGATCAGCA 128
Qy      18 ValAlaLeuAlaGlyPheGluGluAspAsnProIleArgSerValThrGlnArgProAsp 37
Db      129 GCGGCTAATATCGGATCGATGATGATGATGATGATGATGATGATGATGATGATGAT 198
Qy      38 SerIleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
Db      189 GAGGTAGAAGAAATCTGTTTCCCAAGATCTTAGGTCAATCTCGTCACGCTTCTCTCTT 248

```


QY 38 SerileGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
Db 200 GAGGTAGAGAAATCTGTTCCAGATCTTAGTCATCTCTGTCATCGTCTCTTTCGCT 259
QY 58 ArgPheAlaArgArgTyrGlyLysSerTyrGlySerGluGluGluLeuLysLysArgPhe 77
Db 260 CGCTTCACTACCGATATGTTAAAGATATCAAGAGCTGGAGAGATGAAGCTTGCATTC 319
QY 78 GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyr 97
Db 320 TCGATTTTCAAGAGAAATCTTGAATTCAGATCCACCAAGAAAGGCTTATCTTAC 379
QY 98 ThrLeuGlyLysAsnGlnPheAlaAspLeuThrTTPGluGluPheArgThrAsnArgLeu 117
Db 380 AAATCGGTGTAAATCAATTTGCTGATTTGATGCAATGCAAGAGTTTCAAAGGACCAAGCTT 439
QY 118 GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyVal 137
Db 440 GTGCTGCTCAGAACTGCTCTGCCACTTTAAAGGGCAGCCCAAGGTCAAGAGCAGCT 499
QY 138 LeuProValThrArgAspTTPArgGluGlnGlnIleValSerProValLysAspGlnGly 157
Db 500 CTTCTGAAACAAAGAGCTGGAGAGAGATGCTATCGTTAGTCCGGTCAAAGATCAGGA 559
QY 158 SerCysGlySer---TTPThrPheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGln 176
Db 560 GGTGTGGATCTTGTGGACATTCAGCAACTGGAGCTCTTGGAGCAGCTTACCATCAG 619
QY 177 LeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPhe 194
Db 620 GCATTTGGAAGGAATATCTCTCTGAGCAACAGCTTGTGGATTTGCTGGAGCTTTC 679
QY 195 AsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsn 213
Db 680 AATAACTATGTTGCAATGGTGGCTTCTCTCAAGCTTTTGAATACATATAAATCCAAC 739
QY 214 GlyGlyIleAsnThrGluGlnThrTyrProTyrLeuGlyValMetGlyIleCysAsnPhe 233
Db 740 GTGGCTTCGACACAGAGAACTTATCTTACCGTAAAGATGAACCTGCAATTT 799
QY 234 LysGlnGluAsnValGlyValLysValIleAspSerIleAsnIleThrLeuGlyAlaGlu 253
Db 800 TCAGCTGAAACGGTTGGTGTACAAAGTCTCAACTCAGTCAGTCACTTACTCTGGTCTGAA 859
QY 254 AspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGluValVal 273
Db 860 GATGAACATGAAGCATCGGTGGATTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 919
QY 274 LysGlyPheAsnLeuTyrLysLysGlyValTyrSerSerAspThrCysGlyArgAspPro 293
Db 920 CACTCGTTCGGCTTTTACAAGAGTGGAGTTTACACTGATAGTCACTGTGGAGTACTCCA 979
QY 294 MetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIleProTyr 313
Db 980 ATGGATGTGAACACACGGGTTTTTGGCCGTTGGTTATGGAGTTGAAGACGGTGTACCATAT 1039
QY 314 TTPLeuIleLysAsnSerTTPGlyThrAsnTTPGlyAspAsnGlyTyrPheLysMetGlu 333
Db 1040 TGGCTTATTAAGACTCATCGGGAGCGGATTTGGGCGCAAGAGTTACTTCAAGATGAG 1099
QY 334 LeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIleValAla 351
Db 1100 ATGGGGAAGAAATGTTGGTATGCTACATGTCATCATACCCCGTTGGCT 1153

RESULT 13
AF083703
LOCUS AF083703 1355 bp mRNA linear PIN 29-OCT-2002
DEFINITION Arabidopsis thaliana clone spe232 unknown mRNA.
ACCESSION AF083703
VERSION AF083703.1 GI:24417303
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1355)
Stracke,R. and Palme,K.
Signal Peptide Selection derived cDNAs from Arabidopsis thaliana leaves and guard cells
Unpublished
2 (bases 1 to 1355)
Stracke,R. and Palme,K.
Direct Submission
Submitted (12-AUG-1998) Max-Delbrueck-Laboratorium in der Max-Planck-Gesellschaft, Carl-von-Linne-Weg 10, Koeln D-50829, Germany
FEATURES
source
1..1355
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="spe232"
/notes="signal peptide selection derived"
66..1142
/codon_start=1
/product="unknown"
/protein_id="AA060262.1"
/db_xref="GI:24417304"
/translation="MSAKTILSSVLVILVAASAAANIGFDESNPIMVSDGLREVEE SVQIILQSRHVLSPFKTRHYGKYQNVEMKRLRFSIFKENDLRITNKGSLYKL GNQPADLWQEPFRTKLGAAQCSATLKGSHKVTAEALPETKMDREDGTVSPVQOG CGSCWTFTSTGALBAHQAFGKGLSLSEQLVDCAGAFNNGYCGLPSPQAFVIK SNGGLTERXAPYTKGDKTCKSAENVQVNLNVNITLGADELKHAHGLVPRVSLA FEVHSRFLKSGVYTDHSGSTPMDVNVHVLAVGVGVEDGVPTWLIKNSWGDWGDK GYFPMEMGNMCGIATCASYPVVA"
CDS
ORIGIN
Alignment Scores:
Pred. No.: 5 98e-107 Length: 1355
Score: 1279.50 Matches: 240
Percent Similarity: 77.93% Conservative: 39
Best Local Similarity: 67.04% Mismatches: 72
Query Match: 68.61% Indels: 7
DB: 8 Gaps: 4
US-10-087-714-2 (1-352) x AF083703 (1-1355)
QY 1 MetAlaAlaLysLeuLeuPhe-----PheLeuLeuPheLeuValSerAlaLeuSer 17
Db 66 ATGCTCGGAAAAACAATCCTCATCATCAGTAGTCTTGGTGGTCTCTCGCCGATCAGCA 125
QY 18 ValAlaLeuAlaGlyPheGluGluAspAsnProIleArgSerValThrGlnArgProAsp 37
Db 126 GCGGCTAATATCGGATTCGATGAGTCAACCCGATCCGATGCTCCGATGCTCTCCG 185
QY 38 SerileGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
Db 186 GAGGTAGAGAATCTGTTCCAGATCTTAGGTCAATCTCGTCACGCTCTCTTTCGCT 245
QY 58 ArgPheAlaArgArgTyrGlyLysSerTyrGlySerGluGluGluLeuLysLysArgPhe 77
Db 246 CGTTCCTACTCCGATGATGTAAGTATCAGACGTGGAGAGATGAAGCTTCAATTC 305
QY 78 GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyr 97
Db 306 TCGATTTTCAAGGAGAATCTTGAATTCAGATCCACCAAGAAAGGCTTATCTTAC 365
QY 98 ThrLeuGlyLysAsnGlnPheAlaAspLeuThrTTPGluGluPheArgThrAsnArgLeu 117
Db 366 AAATCGGTGTAAATCAATTTGCTGATTTGATGCAAGAGTTTCAAAGGACCAAGCTT 425
QY 118 GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyVal 137
Db 426 GGTGCTGCTCAGAACTGCTCTGCCACTTTAAAGGGCAGCCACCAAGGTCAAGAGCAGCT 485

Qy 138 LeuProValThrArgAspTrpArgGluGlnGlyLeuValSerProValLysAspGlnGly 157
 Db 486 CTTCTCTGAAACAAAGACTGGAGAGATGGTATGCTAGTCCGTCGAAGATCAGGGA 545
 Qy 158 SerCysGlySer---TppThrPheSerThrGlyValAlaLeuAlaValThrGln 176
 Db 546 GGTGTGGATCTGCTGGACATTCAGACACATGGAGCTTGGAGCAGCTTACCATCAG 605
 Qy 177 LeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPhe 194
 Db 606 GCATTTTGGAAAAGGAATATCTCTCTGAGCAACAGCTGTGTGATTTGTCTGAGCTTTC 665
 Qy 195 AsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluCysValLysTyrAsn 213
 Db 666 AATAACATATGTTGCAATGTGGCTTCTCTCAGCCCTTGAATACATAAATCCAC 725
 Qy 214 GlyGlyLeuAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyLeuCysAsnPhe 233
 Db 726 GTGGCTGCACACAGAGAAAGCTTATCTCTATACCGGTAAAGATGAAACCTGCAAAATTT 785
 Qy 234 LysGlnGluValGlyValValValValValValValValValValValValValVal 253
 Db 786 TCAGCTGAAACAGTTGTGTGACAGTCTCTCACTCAGTCAACATTTACTCTGGGTGCTGAA 845
 Qy 254 AspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGluValVal 273
 Db 846 GATGAACCTGAAGCATCGGTGGTGGTATGATCGGCAGTAAGCATAGCAATTTGAGGTATA 905
 Qy 274 LysGlyPheAsnLeuTyrLysGlyValTyrSerSerAspThrCysGlyArgAspPro 293
 Db 906 CACTCGTTCGGCTTTTCAAGAGTGGAGTTTCACTGATAGTCACTGTGGAATATCTCCA 965
 Qy 294 MetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyLysProTyr 313
 Db 966 ATGATGTGAACACCGCGTTTGGCGTGGTATGAGTGTGAAGACGGTGTACCATAT 1025
 Qy 314 TrpLeuLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLysMetGlu 333
 Db 1026 TGGCTTATTAAAGACTCATGGGAGCGGATGGGGGACAAAGGTACTTCAAGATGGAG 1085
 Qy 334 LeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProLysValAla 351
 Db 1086 ATGGGGAAGACATGTGTGATTGTCTACATGTGCATCATACCCCGTGTGGCT 1139

RESULT 14
 AF360273 1370 bp mRNA linear PLN 18-SEP-2002
 LOCUS Arabidopsis thaliana putative cysteine proteinase MALT (At5g60360)
 DEFINITION mRNA, complete cds.
 ACCESSION AF360273
 VERSION AF360273.1 GI:13430721
 KEYWORDS Full cDNA.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 1370)
 Yamada, K., Liu, S.X., Sakano, H., Pham, P.K., Ban, J., Chung, M.K.,
 Goldsmith, A.D., Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L.,
 Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J.,
 Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,
 Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T.,
 Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K.,
 Davis, R.W., Ecker, J.R. and Theologis, A.
 Arabidopsis Full Length cDNA Clones
 Unpublished
 2 (bases 1 to 1370)
 Yamada, K., Liu, S.X., Sakano, H., Pham, P.K., Ban, J., Chung, M.K.,
 Goldsmith, A.D., Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L.,
 Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J.,
 Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,
 Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T.,
 Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K.,
 Davis, R.W., Ecker, J.R. and Theologis, A.

TITLE JOURNAL
 COMMENT
 Submitted (13-MAR-2001) Plant Gene Expression Center, 800 Buchanan
 Street, Albany, CA 94710, USA
 RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN
 Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J.,
 Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
 Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the
 sequencing and annotation of the RAFL cDNAs: Yamada, K., Liu, S.X.,
 Sakano, H., Pham, P.K., Ban, J., Chung, M.K., Goldsmith, A.D.,
 Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L., Chen, H.,
 Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Lam, B., Lin, J.,
 Miranda, M., Nguyen, M., Palm, C.J., Shinn, P., Southwick, A.,
 Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to
 this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC)
 contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis
 genome submitted to Genbank.

FEATURES
 source

1..1370
 /organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/chromosome="5"

/clone="RAFL09-18-G19 (R09257)"

/ecotype="Columbia"

/note="This clone is in a modified pBluescript vector
 (FLC-1) as a BamHI/XhoI insert."

1..1370

/gene="At5g60360"

1..39

/gene="At5g60360"

/misc_difference 1..2

/gene="At5g60360"

/note="compared to genomic sequence"

/replace="tg"

misc_difference 4

/gene="At5g60360"

/note="compared to genomic sequence"

/replace="g"

40..1116

/gene="At5g60360"

/note="contains a papain family cysteine protease domain
 PF00112"

/codon_start=1

/evidence=experimental

/product="putative cysteine proteinase MALT"

/protein_id="AAK25983.1"

/db_xref="GI:13430721"

/translation="MSAKTILSSVVLVAASAAANIGFDESPIRMVSDGLREVLE
 SVSQTIGOSRHVLSFARFTHRYGKYNVEMKLFSGIKENLDLIRSTNKGLSYKL
 GYNGFADLTQCFORTKLGRAONCSATLKGSHKYTEALPETKDWREDGIVSPVKDG
 SCGCWTFSTPTGALEAAHYQAFKGISLSEQLVDGAGANNYCGNGLSQAPFEYIK
 CNGGDTETKATPYTGKDETKCFSAENVGVQVINSVNTLGADELKXAGLVPRVSI
 FEVTHSFLYKSGYVTDHSCGTPDVNHAVLAVGYGVEDGVFWLTKNSWGDWGDGK
 GYFKMEMGKMGCIATCASPVA"

1117..1370

/gene="At5g60360"

1354

/gene="At5g60360"

/note="artifact within poly A tail"

ORIGIN

Alignment Scores:

Pred. No.: 6,066-107 Length: 1370

Score: 1279.50 Matches: 240

Percent Similarity: 77.93% Conservative: 39
Best Local Similarity: 67.04% Mismatches: 72
Query Match: 68.61% Indels: 7
DB: Gaps: 4

US-10-087-714-2 (1-352) x AF360273 (1-1370)

QY 1 MetAlaLeuLeuLeuPhe-----PheLeuLeuPheLeuValSerAlaLeuSer 17
LOCUS : : : : :
Db 40 ATGTTCGCGAAACAATCTCATCATAGTAGTTTGGTGTTCTCGCGCCGCATACGA 99
QY 18 ValAlaLeuAlaGlyPheGluGluAspAenProIleArgSerValThrGlnArgProAsp 37
Db 100 CGCGCTTAATATCGGATTGATGAGTCAAAACCGATCCGAATGGTCTCGATGGTCTCCGG 159
QY 38 SerIleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
Db 160 GAGGTAGAAGAATCTGTTTCCCAGATCTTAGTCAATCTCGTCAAGTCTCTCTTTTCGCT 219
QY 58 ArgPheAlaArgTyrglySerTyrglySerGluGluGluLysLeuLysLeuAspPhe 77
Db 220 CGGTTCACTCACGATATGTTAAAAGATATGAAAGTGGAGGAGGATGAAGCTTCGATT 279
QY 78 GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTy 97
Db 280 TCGATTTTCAAGGAGAATCTTGATTTGATCAGATCCCAACAAGAAAGGCTTATCTTAC 339
QY 98 ThrLeuGlyLysAsnGlnPheAlaSerLeuThrpGluGluPheArgThrAsnArgLeu 117
Db 340 AAACCTCGGTGTAATACTTCTGATTTGATGATGCCAAGAGTTTCAAAGACCACAGCTT 399
QY 118 GlyAlaAlaGlaAsnCysSerAlaThrAlaHisGlyAsnHisAsGlyPheValAspGlyVal 137
Db 400 GTGTGCTGCTCAGACTGCTCTCCACTTTAAAGGCGACCCACAGGTACAGAGCAGCT 459
QY 138 LeuProValThrArgAspTrpArgGluGlnGlyIleValSerProValLysAspGlnGly 157
Db 460 CTTCCTGAAACAAACAACTGGAGAGAAGATGTTATCGTTAGTCGGCTCAAGATCAGGGA 519
QY 158 SerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaThrGln 176
Db 520 GGTTGTGGATCTTGTGGACATTCAGCACTGAGCTCTTGGAGCAGCTTACATCAG 579
QY 177 LeuThrGlySer-----ThrLeuSerGluGlnLeuValAspCysAlaSerAlaPhe 194
Db 580 GCATTTGGAAGAAAGAAATTTCTCTCTGAGCAACAGCTTGTGGATTGCTGGAGCTTTC 639
QY 195 AsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrrValLysTyrrAsn 213
Db 640 AATAACTATGGTTGCAATGGTGGCTTCTCTCAAGCTTTGAAATACATCAATCAATCCAAC 699
QY 214 GlyGlyIleAspThrGluGlnThrTyrrProTyrrLeuGlyValMetGlyIleCysAsnPhe 233
Db 700 GTGTGGCTCGACACAGAGAAGCTTATCTTATACCGGTAAAGTAAACCTGCAAAATTT 759
QY 234 LysGlnGluAsnValGlyValLysValleAspSerIleAsnIleThrLeuGlyValaGlu 253
Db 760 TCAGCTGAAACAGTGTGTGTACAGTCTCACTCACTCAGTCAACATTTACTCTGGTGGTGA 819
QY 254 AspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGluValVal 273
Db 820 GATCAACTAGCATCGGTTTGATTTGGTACGCCAGTACATAGCATTTGAGGTATA 879
QY 274 LysGlyPheAsnLeuTyrrLysLysGlyValTyrrSerSerAspThrCysGlyArgAspPro 293
Db 880 CATCTGTCGGCTTTTACAAAGAGTGGAGTTTCACTGATAGTCACTGTGGAGTACTCCA 939
QY 294 MetAspValAsnHisAlaValLeuAlaValGlyTyrrGlyValGluAspGlyIleProTy 313
Db 940 ATGATGTGAACCAACCGGTTTGGCCGTTGTTGATGAGTTGAAGACGGTGTACCATAT 999
QY 314 TrpLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrrPheLysMetGlu 333
LOCUS : : : : :
5'UTR : : : : :

Db 1000 TGGCTTATTAAAGAACTCATGGGGAGCGGATTTGGGGCGAACAAAGTTACTTCAAGATGGAG 105

QY 334 LeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrrProIleValAla 351
Db 1060 ATGGGGAAGAACATGTGTGGTATTGCTACATGTGATCATACCCCGTGTGGCT 1113

RESULT 15

BT000673 1409 bp mRNA linear PLN 01-OCT-2002

LOCUS Arabidopsis thaliana clone RAFLO7-10-L02 (R10707) putative cysteine proteinase AAPLP (At5g60360) mRNA, complete cds.

DEFINITION BT000673

ACCESSION BT000673.1 GI:23397067

KEYWORDS FLI CDNA.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 1409)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

AUTHORS Yanada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinzaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.

TITLE Arabidopsis Full Length cDNA Clones

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1409)
Yanada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinzaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.

TITLE Direct Submission

JOURNAL Submitted (01-OCT-2002) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA

COMMENT RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAPL cDNAs (RAPL cDNA: 'RIKEN Arabidopsis Full-length cDNA'): Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinzaki,K.

The Salk, Stanford, PGSC (SSP) Consortium members carried out the sequencing and annotation of the RAPL cDNAs: Yanada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yu,G., Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J., Nguyen,M., Palm,C.J., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Davis,R.W., Ecker,J.R. and Theologis,A.

Yanada,K. (SSP/PGEIC) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinzaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGEIC) contributed equally to this work as PIs.

Annotation based on January 2002 version of the Arabidopsis genome submitted to GenBank.

FEATURES
Location/Qualifiers
1..1409
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="RAFLO7-10-L02 (R10707)"
/ecotype="Columbia"
/note="This clone is in a modified pBluescript vector (FLC-1) as a BamHI/XhoI insert."

gene 1..1409
5'UTR /gene="At5g60360" 1..72

Result	No.	Score	Query Match	Length	DB	ID	Description
	1	1819	97.5	1071	9	ADG61376	DNA encod
	2	1310.5	70.3	1442	2	AAQ35111	Enhanced
	3	1300.5	69.7	1407	4	AAQ44802	Cysteine
	4	1300.5	69.7	1407	9	ADB94743	Programme
	5	1279.5	68.6	1358	3	AAC37877	Arabidops
	6	1279.5	68.6	1370	3	AAC48945	Arabidops
	7	1270	68.1	1467	2	AAV82457	Triticum
	8	1265.5	67.9	1077	3	AAU51435	Arabidops

PI Farrell A, John I;
 XX WPI; 1995-131361/17.
 XX
 PT DNA constructs which modify expression of senescence-related genes -
 XX useful to accelerate or inhibit senescence in crop plants.
 XX
 PS Claim 16; Page 55-56; 78pp; English.
 XX
 CC The sequences given in AAT35095-133 are senescence related genes and
 CC fragments which were used in the construct of the invention which
 CC modifies the expression of at least one senescence related gene. Using
 CC these constructs senescence may be inhibited or accelerated in plants
 CC including tomato, lettuce, cabbage, banana, strawberry, wheat, maize,
 CC rice, rape or soybean. Delayed senescence may indirectly prolong the life
 CC of the plant, increase yield, increase protein content of fruits, improve
 CC quality of leafy vegetables, improve tolerance to disease, drought or
 CC other stresses. Increased senescence may more rapidly break down unwanted
 CC plant material and so avoid the use of desiccants on crops. This sequence
 CC represents the senescence-enhanced clone, SEE1, which is also known as
 CC clone p16.4. SEE1 is a cDNA of approx. 1.7 kb encoding a mRNA of approx.
 CC 1.2 kb. The mRNA encoded by SEE1 increases in abundance during maize leaf
 CC senescence. SEE1 shows homology to genes for two thiol proteases, oryzain
 CC gamma from rice and aleurain from barley. The clone SEE1 may therefore
 CC encode a protease. This is supported by the finding that a 1 kb DNA
 CC fragment is amplified by PCR when one of the primers used is derived from
 CC a region which is conserved in a range of thiol proteases. The predicted
 CC amino acid sequence of the protein encoded by SEE1 suggests that the
 CC protein is 360 amino acids long with a mol. wt. of 39 kD. There are
 CC conserved motifs in the amino acid sequence, a putative vacuolar signal,
 CC a Cys active site, a His active site and an Asn active site. The
 CC predominantly hydrophilic protein sequence suggests that the protein is
 CC soluble. SEE1 is deposited as NCIMB 40582
 XX
 SQ Sequence 1442 BP; 314 A; 408 C; 393 G; 327 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 8 418-137 Length: 1442
 Score: 1310.50 Matches: 257
 Percent Similarity: 78.61% Conservative: 26
 Best Local Similarity: 71.39% Mismatches: 68
 Query Match: 70.27% Indels: 9
 DB: 2 Gaps: 6
 US-10-087-714-2 (1-352) x AAT35111 (1-1442)
 QY 1 MetAlaAlaLeuLeuPhePheLeuLeuPheLeuValSerAlaLeuSerValAlaLeu 20
 DB 78 ATGGCCCCACGGCGCTGCTGCTCTCGCGCTGCTGCGCTCGCGCCACCGCGCGCG 137
 QY 21 Ala-----GlyPheGluAspAsnProIleArgSerValThrGlnArgProAspSer 38
 DB 138 GCCAACTCCGGCTCGCGGACTCCAAACCGATCCGCCCGCTCACCGACCGCGCGCTCC 197
 QY 39 ---IleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPhePheAla 57
 DB 198 GCGCTCGAGTCCACCGCTTCGCGCGCTCGCGCCACCGCGCGCGCTCGCTTCGCA 257
 QY 58 ArgPheAlaArgArgTyrGlyLysSerTyrGlySerGluGluIleLysLysArgPhe 77
 DB 258 CGCTTCGCGCTCAGGTACGGCAAGAGTACGAGAGCGCGCGGAGGTCCACAGCGGTT 317
 QY 78 GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyr 97
 DB 318 AGGATCTTCTCCGAGAGCTCCAGCTGCTCGCTCCACCAACCGCAAGGCTCTCTCTAC 377
 QY 98 ThrLeuGlyIleAsnGluPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeu 117
 DB 378 CGCTTCGCGATCAACCGCTTCGCGACATGAGTGGGAGGATTCGCTGCCACCGCGCTC 437
 QY 118 GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPhe-----ValAsp 135
 DB 438 GGTGAGGCCAGAACTGCTCGGCCAGCTTACCGGCAACCAACCGGATCGCGCGCGCC 497

QY 136 GlyValLeuProValThrArgAspTrpArgGluGlnGlyIleValSerProValLysAsp 155
 DB 498 GTTTCGCGTCCCGAGAGCAAGAGATGGAGGAGATGGATTTGTAGCCCGCAGTAAAC 557
 QY 156 GlnGlySerCysGlySer---TyrThrPheSerThrThrGlyAlaLeuGluAlaAlaTyr 174
 DB 558 CAGGCGCACTGTGATCATGCTGAGCTTCCAGCACTACTGGTGCATTTGAGGCTGCATAT 617
 QY 175 ThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSer 192
 DB 618 ACCCAGCACTGGCAAGCCCATCTCTCTCTGAGCAACAGCTTGTGACTGTGGTTT 677
 QY 193 AlaPheAsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLys 211
 DB 678 GCATTCAACATTTCGATGCAACGAGGCGCTTCATCCAGGCGCTTTGANTACATCAA 737
 QY 212 TyrAsnGlyGlyIleAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyIleCys 231
 DB 738 TACAATGGTGGCTTGACACTGAGGAATCTTACCCCTTACCAAGTGTCAATGGAATCTCC 797
 QY 232 AsnPheLysGlnGluAsnValGlyValLysValLysSerIleAsnIleThrLeuGly 251
 DB 798 AAGTTTAAAGAAATGAGAAATGTTGAGTCAAGGTTTTCGACTCGGTTAACATCACCCCTGGT 857
 QY 252 AlaGluAspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGlu 271
 DB 858 GCTGAGGATGAATGAAGGATGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 917
 QY 272 ValValLysGlyPheAsnLeuTyrLysGlyValTyrSerSerAspThrCysGlyArg 291
 DB 918 GTGATCATGCTGTTTCAGGCTGTACAGAGCGGAGTTTACACTAGCCACCATTTGTGAAC 977
 QY 292 AspProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIle 311
 DB 978 ACACCGATGGATGTGAACCAAGCTGTTCTGGCTGTTGGCTGCTGCTGCTGCTGCTGCTG 1037
 QY 312 ProTyrTrpLeuLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLys 331
 DB 1038 CCTACTGCTCATCAAGAACTCATGGCGCGCTGACTGGGGGTGATGAGGGTTACTTCAAG 1097
 QY 332 MetGluLeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIleValAla 351
 DB 1098 ATGGAATGGCAAGAACTGCTGCGGTGTTGCTACGTGTGCTACCTTACCTTATGTGCGCA 1157
 RESULT 3
 AAF44802
 ID AAF44802 standard; cdna; 1407 BP.
 XX
 AC AAF44802;
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE Cysteine protease coding sequence #16.
 XX
 KW Cell death modulator; programmed cell death; PCD; apoptosis;
 XX forestry plant; ss.
 OS Eucalyptus grandis.
 XX
 FN WO200075331-A1.
 XX
 ED 14-DEC-2000.
 XX
 XX 02-JUN-2000; 2000WO-NZ000086.
 PF
 PR 04-JUN-1999; 99US-00325932.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX
 PI Flinn B, Lasham A;
 XX

DR WPI: 2001-061724/07.
 DR P-PSDB; AAB65775.
 XX Novel defender against cell death polynucleotide useful for modulating
 PT programmed cell death pathway and specific development pathways in
 PT forestry plant.
 XX
 XX
 XX
 XX Claim 1; Page 89-90; 142pp; English.
 CC The present invention relates to coding sequences (see AAF44740-P44840
 CC and AAF4843-P44844) and proteins (see AAB65714-B65814) involved in
 CC programmed cell death (PCD; apoptosis). The coding sequences and proteins
 CC of the present invention are useful for modulating a PCD or cell death
 CC pathway and various developmental pathways in a forestry plant, by stably
 CC incorporating one of the present coding sequences into the genome of the
 CC forestry plant, where the coding sequence provides a PCD pathway that is
 CC not present in a native form of the forestry plant
 XX
 XX Sequence 1407 BP; 349 A; 345 C; 377 G; 336 T; 0 U; 0 Other;
 SQ

Alignment Scores:
 Pred. No.: 1,08e-135 Length: 1407
 Score: 1300.50 Matches: 247
 Percent Similarity: 79.77% Conservative: 33
 Best Local Similarity: 70.37% Mismatches: 60
 Query Match: 69.73% Indels: 11
 DB: 4 Gaps: 5

US-10-087-714-2 (1-352) x AAF44802 (1-1407)

QY 9 LeuLeuPheLeuValSerAlaLeuSerValAlaLeuAlaGlyPheGluGluAspAsnPro 28
 DB 91 CTCCTCTCTGCGGCTGCTGCTGCTGCGCGGGCTGAGCTTCGAGGAGTCCACCC 150
 QY 29 IleArgSerValThrGlnArgProAsp-----SerIleGluProAlaIleLeu 44
 DB 151 ATCCGGCTCTC-----CCGACGGGGCTCCGCGACCTCGAGTCTCCATCGTC 201
 QY 45 GlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAlaArgGlyGly 64
 DB 202 CAGATCTGCGCGCGACCCCGCCAGCCCTCTCTTCCGCGCTTCGCCAACAGGTATGG 261
 QY 65 LysSerTyrGlySerGluGluGluIleLysLysArgPheGlyIlePheValGluAsnLeu 84
 DB 262 AAGAGGTACGAGACCGCGGAGGAGATCAAGCTCGGGTTCGAGATCTTCAGGGAGATCTC 321
 QY 85 AlaPheIleArgSerThrAsnArgLysAspLeuSerTyrThrLeuGlyIleAsnGlnPhe 104
 DB 322 AAGTTGATCCGATCCACCAACAGAGGGCTTCGCCCTACACCTCGGTGTCATTAAGTTT 381
 QY 105 AlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAlaAlaGlnAsnCysSer 124
 DB 382 GCTGATTGGAGCTGGGAGGATTTCAGAGGCACAGACTGGGAGCTGCTCAAACTGCTCT 441
 QY 125 AlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProValThrArgAspTrp 144
 DB 442 GCCACCACCAAGGGCAACCAAGCTCACCAGCAAGCTCTCCCGAGATGAAGACTGG 501
 QY 145 ArgGluGlnGlyIleValSerProValLysAspGlnGlySerCysGlySer---TrpThr 163
 DB 502 AGAARAGGGCATGTAAAGCCCAATTAAGATCAGGGCCACTGTGATCTTCTGACT 561
 QY 164 PheSerThrThrGlyAlaLeuGluAlaLysThrGlnLeuThrGlySer-----Thr 181
 DB 562 TTCAGTACCACTGGAGCTCTTGGGCTGCTTATCACCAGCATTCGGGAACAAATCTCT 621
 QY 182 LeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPheGlyCys---Gly 200
 DB 622 CTGCTGTGAGCAGAGCTCGTGACTGCTGGGCTTTCACCACTTTGATGATGTGGT 681
 QY 201 GlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGlyIleAspThrGluGln 220
 DB 682 GGACTGCCATCCCAAGCCTTTGAGTACGTCAAGTACAAAGCTGCGCTTATACCGAGAA 741

221 ThrTyrProTyrLeuGlyValMetGlyIleCysAsnPheLysGlnGluAsnValGlyVal 240
 742 GCATATCTTATACCGCAGTGGATGGTGAATCTCGGCTGATATATGTTGGTGC 801
 241 LysValIleAspSerIleAsnIleThrLeuGlyAlaGluAspGluLeuLysHisAlaVal 260
 802 CAAGTGTCTGACTCTGTAAACATCACCTTGGTCTGAGGATGAACATAAGCATGCAGTT 861
 261 GlyLeuValAspProValSerValAlaPheGluValValLysGlyPheAsnLeuTyrLys 280
 862 GCCTTCGTCGCGCAGTGGATGGTGCATTCAGGTCGTGAAAGACTTCAGATTGTACAAG 921
 281 LysGlyValTyrSerSerAspThrCysGlyArgAspProMetAspValAsnHisAlaVal 300
 922 TCGGCTGTCTACACAGCAGTACATCGGTAGCACTTCCATGGATGTGAACCATGCTGTT 981
 301 LeuAlaValGlyTyrGlyValGluAspGlyIleProTyrTrpLeuIleLysAsnSerTrp 320
 982 CTCGCTGTGTTATGGAGTTGAAGTGGTGTCCGTTCTGGCTCCTCAAGATCTCTGG 1041
 321 GlyThrAsnTrpGlyAspAsnGlyTyrPheLysMetGluLeuGlyLysAsnMetCysGly 340
 1042 GGAGCAGACTGGGTGACACCGGATCTTCAAGATGGAGATGGGAAAGACATGTGTGA 1101
 341 ValAlaThrCysAlaSerTyrProIleValAla 351
 1102 GTCGCTACTTGTGCATCATACCTGTGTGGCC 1134

RESULT 4
 ADB94743
 ID ADB94743 standard; cDNA; 1407 BP.
 XX
 AC ADB94743;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE Programmed cell death pathway cysteine protease cDNA #16.
 XX
 KW Programmed cell death; plant development; plant cell cycle; ATL2; DAD1;
 KW Dnae; lls; lsd1; nucellin-like aspartic protease; annexin; prohibitin;
 KW fen-like protein; rac2; retinoblastoma-related protein; SIN3; TrfID;
 KW TEGF; xylogenic kinase; pur-alpha; cyteine protease; RPPS-like protein;
 KW gp 91 NADPH oxidase subunit; NPR-like protein; BAG-1;
 KW defender against cell death; lethal leaf spot; lesion stimulating death;
 KW seven in absentia; transcription initiation factor;
 KW testis enhanced gene transcript; gene; ss.
 XX
 OS Eucalyptus grandis.
 XX
 PN US2003082724-A1.
 XX
 PD 01-MAY-2003.
 XX
 PF 14-AUG-2002; 2002US-00219220.
 XX
 PR 04-JUN-1999; 99US-00325932.
 XX
 XX (GENE-) GENESIS RES & DEV CORP LTD.
 XX
 XX Plimm B, Lasham A;
 XX
 XX WPI; 2003-786916/74.
 XX
 XX New isolated polynucleotide useful for modulating programmed cell death,
 PT altering the development cycle of plant cells, and subsequently modifying
 PT plant development.
 XX
 XX Claim 1; Page 71-72; 214pp; English.
 XX
 CC The invention describes an isolated polynucleotide (I) comprising a
 CC nucleotide sequence that is one of 145 fully defined sequences of 221-
 CC 3415 base pairs (bp), given in the specification, and/or its complements,

CC reverse complements, reverse sequences, or sequences having 75, 90 or 95
 CC % sequence identity to or that hybridise under stringent hybridisation
 CC conditions to one of the 145 sequences. The methods and compositions of
 CC the present invention to do with (I) are useful for modulating programmed
 CC cell death and thereby altering the development cycle of plant cells, and
 CC altering plant development. This sequence encodes a protein associated
 CC with the programmed cell death pathway.

XX Sequence 1407 BP; 349 A; 345 C; 377 G; 336 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1,08e-135 Length: 1407
 Score: 1300.50 Matches: 247
 Percent Similarity: 79.77% Conservative: 33
 Best Local Similarity: 70.37% Mismatches: 60
 Query Match: 69.73% Indels: 11
 DB: 9 Gaps: 5

US-10-087-714-2 (1-352) x ADB94743 (1-1407)

Qy 9 LeuLeuPheLeuValSerAlaLeuSerValAlaLeuAlaGlyPheGluGluAspAsnPro 28
 Db 91 CTCCTCTCGTCGCGTCGTCGTCGCGCGCGTCGAGCTTCGAGGAGTCCAAACCC 150
 Qy 29 IleArgSerValThrGlnArgProAsp-----SerIleGluProAlaIleLeu 44
 Db 151 ATCCGCTCTTC-----CCGACGGCGGCTCCGCGACCTCGAGCTCCATCGTC 201
 Qy 45 GlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAlaArgArgTyrGly 64
 Db 202 CAGATCGTCGCGCCACCCGCCAGCGCTTCCTCGCCCGCTTCGCCAACAGGTATGG 261
 Qy 65 LysSerTyrGlySerGluGluGluIleLysArgPheGlyIlePheValGluAsnLeu 84
 Db 262 AAGAGTACGAGACCGCGGAGGAGATCAAGCTCGGTTCAGATCTTCAGGGAGATCTC 321
 Qy 85 AlaPheIleArgSerThrAsnArgLysAspLeuSerTyrThrLeuGlyIleAsnGlnPhe 104
 Db 322 AAGTTGATCCGATCCACCAACAGAGGGCTTCGCCCTACACCTCGGTGTCATAAGTTT 381
 Qy 105 AlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAlaAlaGlnAsnCysSer 124
 Db 382 GCTGATTGGAGCTGGGAGGAGTTCAGAGCACAGACTGGAGCTGCTCAAACTGCTCT 441
 Qy 125 AlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProValThrArgAspTyr 144
 Db 442 GCCACCAACGAGGACACCAAGCTCACCAGAGACTCTCCCGAGATGAAGACTGG 501
 Qy 145 ArgGluGlnGlyIleValSerProValLysAspGlnGlySerCysGlySer----TrpThr 163
 Db 502 AGAGAAAAGGCGATTGTAAGCCCAATTAAGATCAGGGGCACCTGTGATCTTCTGACT 561
 Qy 164 PheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGlySer-----Thr 181
 Db 562 TTCAGTACCACTGGAGCTCTTGAGGCTGCTTATCACCAGCAGCTCGGAAACAAATCTCT 621
 Qy 182 LeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPheGlyCys---Gly 200
 Db 622 CTGTCTCAGCAGCAGCTGCTGAGCTGCTGCTGGGCTTCAACAACTTGGATGAGTGT 681
 Qy 201 GlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGlyIleAspThrGluGln 220
 Db 682 GGACTGTCATCCCAAGGCTTGTAGTACGTCAAGTACACCGTGGCTGTATACCGAGAA 741
 Qy 221 ThrTyrProTyrLeuGlyValMetGlyIleCysAsnPheLysGlnGluAsnValGlyVal 240
 Db 742 GCATATCTTATACCGCAGTGGATGGTAGTGCATAATCTTCGCTGATATGTTGTGTC 801
 Qy 241 LysValIleAspSerIleAsnIleThrLeuGlyAlaGluAspGluLeuLysHisAlaVal 260
 Db 802 CAAGTGTCTGACTCTGTAAATCATCATCCTTGGTGGCTGAGGATGAACCTAAGCATGCA 861
 Qy 261 GlyLeuValArgProValSerValAlaPheGluValLysGlyPheAsnLeuTyrLys 280

Db 862 GCCTTCGTCGCGCAGTGCAGTGTGGCACTTCAGAGTTCGAGATTGTACAAG 921
 Qy 281 LysGlyValTyrSerSerAspThrCysGlyArgAspProMetAspValAsnHisAlaVal 300
 Db 922 TCGGTGTCTACACAGGAGTACATCGCGGTAGCACTTCATCGAGTGAACCATGCTGT 981
 Qy 301 LeuAlaValGlyTyrGlyValGluAspGlyIleProTyrTyrLeuIleLysAsnSerTyr 320
 Db 982 CTCGCTGTGTGTTATCGAGTTGAAGATGGTTCGTTTCGTTCTCATCAAGAAATCTCTGG 1041
 Qy 321 GlyThrAsnTrpGlyAspAsnGlyTyrPheLysMetGluLeuGlyLysAsnMetCysGly 340
 Db 1042 GGAGCAGATCGGGTGACCGAGTACTTCAAGTGGAGATGGGAAGAACATGTGTGGA 1101
 Qy 341 ValAlaThrCysAlaSerTyrProIleValAla 351
 Db 1102 GTCGCTACTTGTGCATCATACCTCTGTGTGGCC 1134

RESULT 5
 AAC37877
 ID AAC37877 standard; DNA; 1358 BP.
 XX AAC37877;
 AC AAC37877;
 DT 17-OCT-2000 (first entry)
 XX Arabidopsis thaliana DNA fragment SEQ ID NO: 18989.
 DE Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway; metabolic pathway;
 KW promoter; termination sequence; ss.
 OS Arabidopsis thaliana.
 XX EP1033405-A2.
 XX 06-SEP-2000.
 XX 25-FEB-2000; 2000EP-00301439.
 PR 25-FEB-1999; 99US-0121825P.
 PR 05-MAR-1999; 99US-0123180P.
 PR 09-MAR-1999; 99US-0123548P.
 PR 23-MAR-1999; 99US-0125788P.
 PR 25-MAR-1999; 99US-0126264P.
 PR 29-MAR-1999; 99US-0126785P.
 PR 01-APR-1999; 99US-0127462P.
 PR 06-APR-1999; 99US-0128234P.
 PR 08-APR-1999; 99US-0128714P.
 PR 16-APR-1999; 99US-0129845P.
 PR 19-APR-1999; 99US-0130077P.
 PR 21-APR-1999; 99US-0130449P.
 PR 23-APR-1999; 99US-0130510P.
 PR 23-APR-1999; 99US-0130891P.
 PR 28-APR-1999; 99US-0131449P.
 PR 30-APR-1999; 99US-0132048P.
 PR 30-APR-1999; 99US-0132407P.
 PR 04-MAY-1999; 99US-0132484P.
 PR 05-MAY-1999; 99US-0132485P.
 PR 06-MAY-1999; 99US-0132486P.
 PR 06-MAY-1999; 99US-0132487P.
 PR 07-MAY-1999; 99US-0132863P.
 PR 11-MAY-1999; 99US-0134256P.
 PR 14-MAY-1999; 99US-0134218P.
 PR 14-MAY-1999; 99US-0134219P.
 PR 14-MAY-1999; 99US-0134221P.
 PR 14-MAY-1999; 99US-0134370P.
 PR 18-MAY-1999; 99US-0134768P.
 PR 19-MAY-1999; 99US-0134941P.
 PR 20-MAY-1999; 99US-0135124P.
 PR 21-MAY-1999; 99US-0135353P.
 PR 24-MAY-1999; 99US-0135629P.

PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 08-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.

PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149923P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151086P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.

PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
Alignment Scores:
Pred. No.: 2,33e-133 Length: 1358
Score: 1279.50 Matches: 240
Percent Similarity: 77.93% Conservative: 39
Best Local Similarity: 67.04% Mismatches: 72
Query Match: 68.61% Indels: 7
DB: 3 Gaps: 4

US-10-087-714-2 (1-352) x AAC37877 (1-1358)

QY 1 MetAlaAlaLysLeuLeuPhe-----PheLeuLeuPheLeuValSerAlaLeuSer 17
Db 81 ATGCTGCGAAACCAATCTATCATCTAGTTTGGTGGTCTCTCGTCGCGCATCAGCA 140
QY 18 ValAlaLeuAlaGlyPheGluGluAspAsnProIleArgSerValThrGlnArgProAsp 37
Db 141 GCGGCTAATATCGGATTCGATGAGTCAAAACCGATCCGAATGGTCTCCGATGGTCTCCGG 200
QY 38 SerIleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
Db 201 GAGGTAGAAAGATCTGTTCCAGATCTTAGTCAATCTCTGTCAGTCTCTCTTTCGCT 260
QY 58 ArgPheAlaArgArgTyrglySerTyrglySerGluGluGluIleLysLysArgPhe 77
Db 261 CGCTTCACTACCGATATGTTAAAGATATCAGAACGTCGAGGAGATGAAGCTTCGATTC 320
QY 78 GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTy 97
Db 321 TCGATTTTCAAGGAGAACTTGTATTTGATTCAGATCCACCAACAGAAAGGTTATCTTAC 380
QY 98 ThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeu 117
Db 381 AAATCGGTGTTATCAATTTGCTGATTTGATGAGTGGCAAGATTTCAAAGACCAAGCTT 440
QY 118 GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyVal 137
Db 441 GGTGTCTGTCTGAGAACTGTCTGCCACTTAAAGGCGCAGCCCAAGGTCACAGAAGCAGCT 500
QY 138 LeuProValThrArgAspTrpArgGluGlnGlyIleValSerProValLysAspGlnGly 157
Db 501 CTTCTCTGAAACAAAGACTGGAGAGAACTGGTATCGTTAGTCCGGTCAAAGATCAGGA 560
QY 158 SerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaIleTyThrGln 176
Db 561 GGTGTGGATCTTGTGGACATTCAGCAAACTGGAGCTCTTGAGGCGAGCTTACCATCAG 620
QY 177 LeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPhe 194
Db 621 GCATTTGGAAAGGAATATCTCTCTGAGCAACAGCTTGTGGATTTGCTGGAGCTTTC 680
QY 195 AsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyThrValLysTyAsn 213
Db 681 AATAACTATGTTGCAATGGTGGCCTCTCTCTCAAGCCTTTGAATACATATAAATCCAAC 740
QY 214 GlyGlyIleAsnThrGluGlnThrTyProTyThrLeuGlyValMetGlyIleCysAsnPhe 233
Db 741 GGTGGCTTCGACAGAGAACTTATCTTATACCGGTAAAGATGAACCTGCAATTT 800
QY 234 LysGlnGluAsnValGlyValLysValIleAspSerIleAsnIleThrLeuGlyValGlu 253
Db 801 TCAGCTGAAACGTTGGTGTACAACTCTCAACTCAGTCAACATTTACTCTGGTGTCTGAA 860
QY 254 AspGluLeuLysPheHisAlaValGlyLeuValArgProValSerValAlaPheGluValVal 273
Db 861 GATGAACGAAACATGCGGTTGGATTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 920
QY 274 LysGlyPheAsnLeuTyrlsLysGlyValTyrlsSerAspThrCysGlyArgAspPro 293
Db 921 CACTCGTTCGCGCTTACAAAGAGTGGAGTTTACACTGTATGATCTACTGTGGAGTACTCCA 980

QY 294 MetAspValAsnHisAlaValLeuAlaValGlyTyThrGlyValGluAspGlyIleProTy 313
Db 981 ATGGAATGTAACACCGCGTTTGGCGTTTGGTTATGGAGTTGAAGACGGTGTACCATAT 1040
QY 314 TrpLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyThrPheLysMetGlu 333
Db 1041 TGGCTTATTAAAGAACTCATGGGAGCGGATTCGGGCGACAAAGGTTACTTCAAGATGGAG 1100
QY 334 LeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyProIleValAla 351
Db 1101 ATGGGAAGAACAATGTGTGTATTGTCTATGTCATCATACCCCGTTGTGGCT 1154

RESULT 6
AAC48945
ID AAC48945 standard; DNA; 1370 BP.
XX
AC AAC48945;
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 59352.
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 23-MAR-1999; 99US-0126264P.
PR 23-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 23-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.

PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 24-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-0140921P.
PR 28-JUN-1999; 99US-0141287P.
PR 30-JUN-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 21-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145132P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 27-JUL-1999; 99US-0145951P.
PR 28-JUL-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.

PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149388P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 13-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160985P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:
Pred. No.: 2,368-133
Score: 1279.50
Percent Similarity: 77.93%

Length: 1370
Matches: 240
Conservative: 39

Best Local Similarity: 67.04% Mismatches: 72
Query Match: 68.61% Indels: 7
DB: 3 Gaps: 4

US-10-087-714-2 (1-352) x AAC48945 (1-1370)

```
QY 1 MetAlaAlaLysLeuLeuPhe-----PheLeuLeuPheLeuValSerAlaLeuSer 17
DB 80 ATGCTTCGGAACAACTCATCATCATGTTTGGTGTCTCGTCGCGCATCAGCA 139
QY 18 ValAlaLeuAlaGlyPheGluGluAspAsnProIleArgSerValThrGlnArgProAsp 37
DB 140 GCGGCTAATATCGGATTCGATGAGTCAAAACCGATCGAATGGTCTCCGATGGTCTCCGG 199
QY 38 SerIleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
DB 200 GAGGTAGAGAAATCTGTTCCAGATCTTAGTCAATCTCGTCACTTCTCTCTCTCTCGCT 259
QY 58 ArgPheAlaArgArgTyrGlyLysSerTyrGlySerGluGluGluIleLysLysArgPhe 77
DB 260 CGCTTCACCTACCGATATGTTAAAGATATCAGAACGTGGAGGAGATGAAGCTTCGATTC 319
QY 78 GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyr 97
DB 320 TCGATTTTCAGGAGAAATCTTGATTTGATCAGATCCACCAACAAAGAAAGGCTTATCTTAC 379
QY 98 ThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeu 117
DB 380 AAATCGGTGTATCAATTTGCTGATTTGATGCGCAAGAGATTTCAAGACCAAGCTT 439
QY 118 GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyVal 137
DB 440 GGTGCTGCTCAGAACTGCTCGCACTTTAAAGGCGCAGCCACAGGTTCACAGAACAGCT 499
QY 138 LeuProValThrArgAspTrpArgGlnGlyIleValSerProValLysAspGlnGly 157
DB 500 CTTCCTGAAACAAAGATCGAGAGAGATGATCGTTAGTTCGGTCAAAAGATCAGGGA 559
QY 158 SerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGln 176
DB 560 GGTGCTGATCTTGCTGACATTCAGCACTGAGCTCTTGAGCGAGCTTACCATCAG 619
QY 177 LeuThrGlySer-----ThrLeuSerGlnGlnLeuValAspCysAlaSerAlaPhe 194
DB 620 GCATTTGAAAGAAAGAAATTTCTCTCTCAGCAACAGCTTGTGGATTTGCTGGAGCTTC 679
QY 195 AsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsn 213
DB 680 AATACTATGTTGCAATGGTGGCTTCTCTCAGGCTTTGAATACATCAATCCAACT 739
QY 214 GlyIleAspThrGlnGlnThrTrpProTyrLeuGlyValMetGlyIleCysAsnPhe 233
DB 740 GGTGCTCGACACAGAGAAAGCTTATCTCTATACCGGTAAAGATGAACCTGCAAAATTT 799
QY 234 LysGlnGluAsnValGlyValLysValIleAspSerIleAsnIleThrLeuGlyAlaGlu 253
DB 800 TCAGCTGAAACGTTGGTGTACAGTCTCACTCACTCACTCACTCACTCACTCACTCACT 859
QY 254 AspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGluValVal 273
DB 860 GATGAATGAACATCGGTTGGATTTGATCGGCCAGTAAGCATAGCATATTGAGGTTATA 919
QY 274 LysGlyPheAsnLeuTyrLysGlyValTyrSerSerAspThrCysGlyArgAspPro 293
DB 920 CACTCGTTCGCGCTTTTACAAGAGTGGAGTTTACACTGATAGTACTGCTGGAGTACTCCA 979
QY 294 MetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIleProTyr 313
DB 980 ATGGATGTGAACACCGGTTTGGCCGTTGGTTATGAGTTGAAGACGGTGTACCATAT 1039
QY 314 TrpLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLysMetGlu 333
DB 1040 TGGCTTATTAAAGAACTCATGGGAGCGGATGGGGCGCAAAAGGTTACTTCAAGATGAG 1099
```

```
QY 334 LeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIleValAla 351
DB 1100 ATGGGGAAGAACATGTGTGTATTGCTATCATGTGTCATACCCCGTTGTGGCT 1153
```

RESULT 7

AAV82457
ID AAV82457 standard; DNA; 1467 BP.

XX AAV82457;

DT 16-MAR-1999 (first entry)

DE Triticum sp. cysteine proteinase #2 encoding DNA.

XX Triticum; wheat seed; cysteine proteinase; gluten; baking; ds.

OS Triticum sp.

XX Key Location/Qualifiers
XX CDS 47..1144
XX /*tag= a

XX JF10327886-A.

XX 15-DEC-1998.

XX 27-MAR-1998; 98JP-00098140.

XX 31-MAR-1997; 97JP-00114946.

XX (SHOS) SHOWA SANGYO CO.

XX WPI; 1999-109255/10.

XX P-PSDB; AAW89557.

XX New DNA coding cysteine proteinase originating from wheat seed - useful

XX for improving gluten for use in bakery process.

XX Claim 2; Page 16-18; 29pp; Japanese.

XX The present sequence encodes a cysteine proteinase isolated from wheat

XX seed (Triticum sp.). The cysteine proteinase is useful for improving

XX gluten for use in the bakery process

XX Sequence 1467 BP; 348 A; 423 C; 389 G; 305 T; 0 U; 2 Other;

Alignment Scores:
Pred. No.: 3 04e-132 Length: 1467
Score: 1270.00 Matches: 246
Percent Similarity: 78.12% Conservative: 36
Best Local Similarity: 68.14% Mismatches: 67
Query Match: 68.10% Indels: 12
DB: 2 Gaps: 6

US-10-087-714-2 (1-352) x AAV82457 (1-1467)

QY 3 AlaLysLeuPhePheLeuLeuValSerAlaLeuSerValAlaLeuAla--- 21

DB 56 GCGCGATCTCTCTCTCGGCTCGGCGCCACCGCCACCGCTCGCGCTCC 115

QY 22 -----GlyPheGluGluAspAsnProIleArgSerValThrGlnArgPro 36

DB 116 TCCTCTCTCTCTCTCTCTCTCTCAACCGATCCGCGCCGCTCACCAGCGCGCC 175

QY 37 AspSer---IleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHis 55

DB 176 GCTTCACGCTCGAGTCCCGGCTCTCGCGCGCTCGGCGCCACCGCCACCGCTCCGC 235

QY 56 PheAlaArgPheAlaArgTyrGlyLysSerTyrGlySerGluGluGluIleLysLys 75

DB 236 TTCGCGGCTTCCGCTCAGGTACGCAAGAGTACAGAGCGCGCGGAGTCCGATG 295

QY 76 ArgPheGlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgIysAspLeu 95
DB 296 CGGTTCCGGATCTTCGAGAGCTCGAGGAGTGCCTCCACCAACCGGAAGGCGCTA 355
QY 96 SerTyrThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsn 115
DB 356 TCTACCGCTCGGATCAACCGCTCTCGGACATGAGTGGAGGAGTTCAGGCGACC 415
QY 116 ArgLeuGlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAsp 135
DB 416 CGGCTCGGCGCGCGAGACCTGCTCGGCACGCTCGCGGCAACCACTGATGCGGAC 475
QY 136 ---GlyValLeuProValThrArgAspTrpArgGluGlnGlyIleValSerProValIys 154
DB 476 GCGCGCGCTCCGGAGACCAAGACATGGAGGAGGATGGATCGTTAGCCCGTAAA 535
QY 155 AspGlnGlySerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAla 173
DB 536 GACCACTCCATTGCGGCTCTCGGATCGGACGTTTCAGCACTACTGTCGCACTTGAGGACGA 595
QY 174 TyrThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAla 191
DB 596 TATACTCAGGCACCTGGAAGAACATCTCTCTTTCTGACCAACACTGCTGGTGACTGTGCC 655
QY 192 SerAlaPheAsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrVal 210
DB 656 GCGCGATTCAATAATTCGGATCGAGCGAGGCGCTTCATCCAGGCATTTGAGTACATC 715
QY 211 LysTyrAsnGlyGlyIleAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyIle 230
DB 716 AAATACATGGAGGATCGATACCGAGGAGTCTACCCCTTCAAGGGTGTCTAATGGCGTC 775
QY 231 CysAsnPheLysGlnGluAsnValGlyValIysValIleAspSerIleAsnIleThrLeu 250
DB 776 TGCACTTACAGCTGAAATGCGGTAGTTTCAAGTTTGGACTCTGTTAAACATCACACTG 835
QY 251 GlyAlaGluAspGluLeuIysHisAlaValGlyLeuValArgProValSerValAlaPhe 270
DB 836 ATGCTGGAGGAGAACTGAAACCGCGCTGGGTGGTTGCGCCAGTGTGCTTGCCTTT 895
QY 271 GluValValIysGlyPheAsnLeuTyrLysLysGlyValTyrSerSerAspThrCysGly 290
DB 896 GAGTGATCAACGGTTTCAGGAGTACAGAGCGGAGTTTACATAGCGACCATTTGCGGT 955
QY 291 ArgAspProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGly 310
DB 956 ACTACACCTGATGACGTGAACCAACCGCGCTTGGCGGTGCGGTACCGGTGGAGAACGCG 1015
QY 311 IleProTyrTrpLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPhe 330
DB 1016 GTCCCGATTGGTCTCATCAAGAACTCTGCGGCGCGACTGGGTGACAACTGTTACTTC 1075
QY 331 LysMetGluLeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIleVal 350
DB 1076 AAGATGGAATGGCAAGAACATGT 1135
QY 351 Ala 351
DB 1136 GCG 1138
RESULT 8
AAC51435
ID AAC51435 standard; DNA; 1077 BP.
XX
AC AAC51435;
XX
DT 18-OCT-2000 (first entry)
DE
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 68510.
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.

XX OS
XX FN Arabidopsis thaliana.
XX PN EPI033405-A2.
XX PD
XX XX 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-00301439.
XX XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.

QY 98 ThrLeuGlyLeuAsnGlnPheAlaSerLeuThrTrpGluGluPheArgThrAsnArgLeu 117
DB 301 AAATCTCTCTTAATCAGTTTGTGACTTGTGACATGCGAAGATTTCAAAGATCAAGCTT 360
QY 118 GlyAlaAlaGlnAsnGlnSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyVal 137
DB 361 GGNAGTCTCTCAAACTGTTCTGCTACTTTAAAGGGTAGCCACAAGACTCACTGAAGCTACA 420
QY 138 LeuProValThrArgAspTrpArgGluGlnGlyLeuValSerProValIleAspGlnGly 157
DB 421 GTTCCAGACACAAAGATTGGAGAGAGATGTTTGTAGCCCTGTGAAAGAACAGGGA 480
QY 158 SerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaLysThrGln 176
DB 481 CATTTGGATCTCTTGGACATTTAGCACACTGGAGCTCTTGAAGAGCTTACCATCAA 540
QY 177 LeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPhe 194
DB 541 GCATTTGGAAGGAAGATATCTTTGTCGAGCAACAGCTTGTGGATTGTGCTGCTACTTC 600
QY 195 AsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValIleSerThr 213
DB 601 AATAACTTTGGTTGTTCATGGTGACTTCTTCAAGCCCTTTGAATACATTAATAATACAC 660
QY 214 GlyGlyLeuAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyLeuCysAsnPhe 233
DB 661 GTTGGGCTCGACACGAGGAGGCTTATCTTACCCGGAAGAGCGGTGGCTGCAATTT 720
QY 234 LysGlnGluAsnValGlyValIleValIleAspSerIleAsnIleThrLeuGlyAlaGlu 253
DB 721 TCAGCGAAACATCGGTGTACAAAGTCGCTGACTCTGTCAACATTACCTCGGTGTCAGAA 780
QY 254 AspGluLeuLeuHisAlaValGlyLeuValArgProValSerValAlaPheGluValVal 273
DB 781 GATGAATGAGACACGCGGTGGTTGGTAAGCCAGTGTGGCGTTTGAGGTGTA 840
QY 274 LysGlyPheAsnLeuTyrIleValIleValIleValIleValIleValIleValIleVal 293
DB 841 CATGAATTCAGGTTTATAGAAGGAGGAGTTTACTAGCAATACATGTGTGTAACACTCA 900
QY 294 MetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIleProTyr 313
DB 901 ATGATGTAAACCATCGAGTGTAGCAGTGTGTATGGAGTTGAAGACGACGCTCCGCTAC 960
QY 314 TrpLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLysMetGlu 333
DB 961 TGGCTTTATAAGAACTCATGGGAGGTGAATGGGAGACAAATGGCTACTTCAAGATGGAA 1020
QY 334 LeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIleValAla 351
DB 1021 ATGGGAAGAACATGTGTGTGTGTGACATGTTCATCGTATCCGTTGTAGCC 1074
RESULT 9
ID ABN98252/c
XX ID ABN98252 standard; DNA; 1282 BP.
AC ABN98252;
XX
DT 01-AUG-2002 (first entry)
XX
DE Arabidopsis thaliana expressed polynucleotide SEQ ID NO 20.
XX
KW Arabidopsis thaliana; plant; insecticide; fungicide; transgenic; stress;
KW disease; crop; thale cress; tolerance factor; insect; pathogen;
KW nutrition; ds.
XX
OS Arabidopsis thaliana.
XX
PN US2002023281-A1.
XX
PD 21-FEB-2002.
XX

26-JAN-2001; 2001US-00770445.
27-JAN-2000; 2000US-0178472P.
(GORL// GORLACH J.
(ANY// AN Y.
(HAMI// HAMILTON C M.
(PRIC// PRICE J L.
(RAIN// RAINES T M.
(YUY// YU Y.
(RAME// RAMEAKA J G.
(PAGE// PAGE A.
(MATH// MATHW A V.
(LEDF// LEDFORD B L.
(WOES// WOESSNER J P.
(HAAS// HAAS W D.
(GARC// GARCIA C A.
(KRIC// KRICKER M.
(SLAT// SLATER T.
(DAVI// DAVIS K R.
(ALLE// ALLEN K.
(HOFF// HOFFMAN N.
(HURB// HUREAN P.
Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
Hurban P;
WPI: 2002-403163/43.
New Arabidopsis thaliana nucleic acid for identifying homologous genes,
producing compositions that modulate the expression or function of its
encoded protein, and mapping functional regions of protein.
Claim 1; SEQ ID NO 20; 49pp + Sequence Listing; English.
The invention relates to an Arabidopsis thaliana nucleic acid (I)
comprising a sequence capable of hybridising under stringent conditions
to a sequence selected from any one of 999 sequences (ABN98233-ABN99231),
given in the specification or its fragment. A polypeptide (II) encoded by
(I), a transgenic plant (III) comprising an exogenous nucleic acid or a
genetically modified cell (IV) comprising an exogenous nucleic acid, is
useful for screening a candidate agent for its biological effect. (I) is
useful in identifying homologous or related genes, in producing
compositions that modulate the expression or function of its encoded
protein, mapping functional regions of the protein and in studying
associated physiological pathways. (I) is also useful for the genetic
manipulation of cells, particularly plant cells. (I) is also useful in
screening assays of various plant strains to determine the strains that
are best capable of withstanding a particular disease or environmental
stress. (II) and (III) are useful for screening of biologically active
agents, e.g. fungicides, insecticides, etc., for elucidating biochemical
pathways. The screened agents are useful in improved methods of treating
crops to prevent or treat disease. (II) are also useful in screening
programs to identify agents that mimic or enhance the action of tolerance
factors. Such agents are useful in improved methods of treating crops to
enhance their tolerance to environmental stress. (I) is also useful for
enhancing or inhibiting production of a biosynthetic product in a plant.
(III) is useful for identifying other mediators that may induce
expression of proteins of interest, for establishing the extent to which
any specific insect and/or pathogen is responsible for damage to a
particular plant, for identifying other mediators that enhance or induce
tolerance to environmental stress, for identifying factors involved in
biosynthetic pathways of nutritional, commercial, or medicinal value
for identifying productions of nutritional, commercial or medicinal
value. (IV) is useful in the study of genetic function and regulation,
for alteration of the cellular metabolism and for screening compounds
that may affect the biological function of the gene or gene products.
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from USPTO
at seqdata.uspto.gov/sequence.html?DocID=999909770445

Sequence 1282 BP; 367 A; 314 C; 234 G; 367 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.53e-131 Length: 1282
 Score: 1263.00 Matches: 235
 Percent Similarity: 78.39% Conservative: 37
 Best Local Similarity: 67.72% Mismatches: 71
 Query Match: 67.72% Indels: 4
 DB: 6 Gaps: 3

US-10-087-714-2 (1-352) x ABN98252 (1-1282)

Qy 9 LeuLeuPheLeuValSerAlaLeuSerValAlaLeuAlaGlyPheGluGluAsnPro 28
 Db 1252 CTATTGATTCTCTTTGGCGGCGGAGTAAAGATCGGATTCGATGATCAATCCA 1193
 Qy 29 IleArgSerValThrGlnArgProAspSerIleGluProAlaIleLeuGlyValIleuGly 48
 Db 1192 ATCAAATGGTCTCCGCAATCTCCATGAGCTTGAAGACACCGTCTGCATGCTTAGGA 1133
 Qy 49 SerCysArgHisAlaPheHisPheAlaArgPheAlaArgAIGTyrGlyLysSerTyrGly 68
 Db 1132 CAGTCCCGTCATGTTCTATCTCTTCTGCTCACTACAGTATGGGAAGAAGTATCAG 1073
 Qy 69 SerGluGluGluIleLysLysArgPheGlyIlePheValGluAsnLeuAlaPheIleArg 88
 Db 1072 AGTGTGGAGGAGATGAAGTAAAGTCTCTGTTTAAAGGAGATCTTGTATTAATCAGA 1013
 Qy 89 SerThrAsnArgLysAspLeuSerTyrThrIleuGlyIleAsnGlnPheAlaAspLeuThr 108
 Db 1012 TCACATAAAGAAAGGCTTATCTTATAAATCTCTCTTAAATCACTTGTCTGACTTGAC 953
 Qy 109 TrpGluGluPheArgThrAsnArgLeuGlyAlaAlaGlnAsnCysSerAlaThrAlaHis 128
 Db 952 TGGCAAGAGTTTCAAGATACAGCTTGGAGCTGCTCAAACTGTCTGCTACTTTAAAG 893
 Qy 129 GlyAsnHisArgPheValAspGlyValLeuProValThrArgAspTyrArgGluGlnGly 148
 Db 892 CGTACACGAGATCACTACAGTCTCAGACACAAAAGATTGGAGAGAAGTGGT 833
 Qy 149 IleValSerProValLysAspGlnGlySerCysGlySer---TyrThrPheSerThrThr 167
 Db 832 ATTGTAGCCCTGTGAAGACAGGACATTTGTGGATCTTTGGACATTTAGCACAACT 773
 Qy 168 GlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGlySer-----ThrLeuSerGluGln 185
 Db 772 GGAGCTCTTGAAGCAGCTTACCATCAAGCATTTGGAAAAGGAATATCTTCTCCGAGCAA 713
 Qy 186 GlnLeuValAspCysAlaSerAlaPheAsnAsnGlyCys--GlyGlyLeuProSer 204
 Db 712 CAGCTGTGGATGTGCTGCTACTTCTCAATTAATCTTGTGTCTCATGTGGACTTCTCTCT 653
 Qy 205 GlnAlaPheGluTyrValLysTyrAsnGlyGlyIleAspThrGluGlnThrTyrProTyr 224
 Db 652 CAAGCTTTGAATACATTAAATACAAACGCTGGCTCGACACGGAGGAGGCTTATCTTAC 593
 Qy 225 LeuGlyValMetGlyLysCysAsnPheLysGlnGluAsnValGlyValLysValIleAsp 244
 Db 592 ACCGAAAGACGGTGGCTGCAAAATTTACGCAAAACATCGGTGTACAGTCCGCTGAC 533
 Qy 245 SerIleAsnIleThrLeuGlyValAlaGluAspGluLeuLysHisAlaValGlyLeuValArg 264
 Db 532 TCTGTCAACATTACCTGGTGCAGAGATGAATGAATGAATGAATGAATGAATGAATGAAT 473
 Qy 265 ProValSerValAlaPheGluValValLysGlyPheAsnLeuTyrLysGlyValTyr 284
 Db 472 CCAGTGTGGTGGCTTTGAGGTTGTATCAATTAATTAATTAATTAATTAATTAATTAAT 413
 Qy 285 SerSerAspThrCysGlyArgAspProMetAspValAsnHisAlaValLeuAlaValGly 304
 Db 412 ACTAGCATATCATGTGGTAACACTCAATGATGTAAACCATGATGATGATGATGATGAT 353
 Qy 305 TyrGlyValGluAspGlyIleProTyrTrpLeuIleLysAsnSerTrpGlyThrAsnTrp 324

Db 352 TATGGAGTTGAAGACGACGTCCTGCTACTGCTTATAAGACTCATGGGAGGTGAATGG 293
 Qy 325 GlyAspAsnGlyTyrPheLysMetGluLeuGlyLysAsnMetCysGlyValAlaThrCys 344
 Db 292 GGAGACAAATGGCTACTTCAAGATGGAAATGGGAAGAACATGTGTGTGTTGCAACATGT 233
 Qy 345 AlaSerTyrProIleValAla 351
 Db 232 TCATCGTATCCGGTTGTAGCC 212
 RESULT 10
 AAC40628
 ID AAC40628 standard; DNA; 1363 BP.
 AC AAC40628;
 XX 17-OCT-2000 (first entry)
 XX Arabidopsis thaliana DNA fragment SEQ ID NO: 28994.
 DE Arabidopsis thaliana
 KW Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway; metabolic pathway;
 KW promoter; termination sequence; ss.
 OS Arabidopsis thaliana.
 XX EP1033405-A2.
 PN 06-SEP-2000.
 PD 25-FEB-2000; 2000EP-00301439.
 PF 25-FEB-1999; 99US-0121825P.
 PR 05-MAR-1999; 99US-0123180P.
 PR 03-MAR-1999; 99US-0123548P.
 PR 23-MAR-1999; 99US-0125788P.
 PR 25-MAR-1999; 99US-0126264P.
 PR 29-MAR-1999; 99US-0126785P.
 PR 01-APR-1999; 99US-0127462P.
 PR 06-APR-1999; 99US-0128234P.
 PR 08-APR-1999; 99US-0128714P.
 PR 16-APR-1999; 99US-0129845P.
 PR 13-APR-1999; 99US-0130077P.
 PR 21-APR-1999; 99US-0130449P.
 PR 23-APR-1999; 99US-0130510P.
 PR 28-APR-1999; 99US-0130891P.
 PR 30-APR-1999; 99US-0131449P.
 PR 30-APR-1999; 99US-0132048P.
 PR 04-MAY-1999; 99US-0132407P.
 PR 05-MAY-1999; 99US-0132484P.
 PR 05-MAY-1999; 99US-0132485P.
 PR 06-MAY-1999; 99US-0132486P.
 PR 07-MAY-1999; 99US-0132487P.
 PR 11-MAY-1999; 99US-0134256P.
 PR 14-MAY-1999; 99US-0134218P.
 PR 14-MAY-1999; 99US-0134219P.
 PR 14-MAY-1999; 99US-0134221P.
 PR 14-MAY-1999; 99US-0134370P.
 PR 18-MAY-1999; 99US-0134768P.
 PR 19-MAY-1999; 99US-0134941P.
 PR 21-MAY-1999; 99US-0135124P.
 PR 21-MAY-1999; 99US-0135353P.
 PR 24-MAY-1999; 99US-0135629P.
 PR 25-MAY-1999; 99US-0136021P.
 PR 27-MAY-1999; 99US-0136392P.
 PR 28-MAY-1999; 99US-0136782P.
 PR 01-JUN-1999; 99US-0137222P.
 PR 03-JUN-1999; 99US-0137528P.
 PR 04-JUN-1999; 99US-0137502P.
 PR 07-JUN-1999; 99US-0137724P.
 PR 08-JUN-1999; 99US-0138094P.

PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 17-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 24-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-0140981P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143342P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 18-JUL-1999; 99US-0144325P.
PR 18-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145119P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 28-JUL-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147152P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.

PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0158055P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:
Pred. No.: 4, 12e-131
Score: 1259.50
Percent Similarity: 77.44%
Best Local Similarity: 66.57%

Length:
Matches: 1363
Conservative: 239
Mismatch: 39
Mismatches: 73

Query Match: 67.53% Indels: 8
DB: 3 Gaps: 4
US-10-087-714-2 (1-352) x AAC40628 (1-1363)

QY 1 MetAlaAlaLysLeuLeuPhe-----PheLeuLeuPheLeuValSerAlaLeuSer 17
DB 82 ATGCTCGAAGAACAACTCATCATAGTAGTTTGGTGGTTCGTCGCCGCATCAGCA 141

QY 18 ValAlaLeuAlaGlyPheGluGluAspAsnProIleArgSerValThrGlnArgProAsp 37
DB 142 GCGGCTAATATCGGATTCGATGAGTCAAAACCCGATCCGAATGGTCTCCGATGGTCTCCG 201

QY 38 SerIleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
DB 202 GAGTAGAAGAAATCGTTTCCAGATCTTAGTCAATCTCGTCAGCTTCTCTCTTCGCT 261

QY 58 ArgPheAlaArgAArgTyrGlyLysSerTyrGlySerGluGluGluIleLysLysArgPhe 77
DB 262 CGCTTCACTCACCGATATGGTAAAGAGTATCAGAACGTGGAGGAGATGAAGCTTCGATT 321

QY 78 GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyr 97
DB 322 TCGATTTCGAAGGAAGATCTTGATTGATCAGATCCACCAACAGAAAGGCTTATCTTAC 381

QY 98 ThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeu 117
DB 382 AAACCTGGTGTAAATCAATTTGCTGATTGACATGGCAAGAGTTTCAAAGGACCAAGCTT 441

QY 118 GlyAlaAlaGlnAsnCysSerAlaAlaThrAlaHisGlyAsnHisArgPheValAspGlyVal 137
DB 442 GGTCTCTCTCAGAACTGCTCGCCACTTTAAAGGGCAGCCACAGGTCACAGAAGCAGCT 501

QY 138 LeuProValThrArgAspTrpArgGluGlnGlyIleValSerProValLysAspGlnGly 157
DB 502 CTTCCTGAAACAAAGAACTCGAGAGAAGATGGTATCGTTAGTCGGTCAAAGATCAGGA 561

QY 158 SerCysGlySer----TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGln 176
DB 562 GGTGTGGATCTTCCTCGACATTCAGACAACTGGAGCTCTTGAGGCAGCTTACCATCAG 621

QY 177 LeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPhe 194
DB 622 GCATTTGGAAAAGGAATATCTCTCTCAGCAACAGCTTGTGGATGTGCTGGAGCTTTC 681

QY 195 AsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsn 213
DB 682 AATAACTATGTTTCCAATGGTGGCTCTCTCTCAAGCCTTTGGAATACATAATAATCCAAC 741

QY 214 GlyGlyIleAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyIleCysAsnPhe 233
DB 742 GTGGCCCTCGACACAGAGAAAGCTTATCCCTATACCGTGAAGATGAACCTGCAATTT 801

QY 234 LysGlnGluAsnValGlyValLysValIleAspSerIleAsnIleThrLeuGlyValGlu 253
DB 802 TCAGCTGAAACAGTTGGTGTACAAGTCTCAACTCAGTCAACATTAATCTCTGGGTGCTGAA 861

QY 254 AspGluLeuLysHisAlaValI-GlyLeuValArgProValSerValAlaPheGluValVa 273
DB 862 GATGAACCTGAAGCATGCGGTTGAGATTGGTAGCCGACAGTAAAGCATGATTTGAGGTTAT 921

QY 273 LysGlyPheAsnLeuTyrLysLysGlyValTyrSerSerAspThrCysGlyIleAspPr 293
DB 922 ACACCTGTTCCGGCTTTACAAGAGTGGAGTTTACACTGATAGTCACGTGGAAGTACTCC 981

QY 293 oMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIleProTy 313
DB 982 AATGGATGTGAACCAACCGGGTTTGGCCGTTGGTATGGAGTTGAAGACGGTGTACCAT 1041

QY 313 rTrpLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLysMetGl 333
DB 1042 TTGCTTATTAAAGAACTCATCGGAGCGGATTTGGGGCAGCAAAAGGTTACTTCAAGATGGA 1101

Db 20 TCACGTCCTCGCTCGTATTGTTCTCGTCGCCGCCCTTTTCGCTACTGCACTTGCCTG 79
QY 23 -----PheGluGluAspAsnProIleArgSerValThrGlnArgProAspSerIle 39
Db 80 CGCGGACCTTCGCTGATAAGATCCGATCAGCAAGTCGTA---TTTCGCGATGAGCTG 136
QY 40 GluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPhe 59
Db 137 GAGACGGGATTCCTCAAGTCGTGGCCAGACTCGCAGTGTCTCTCTCTCGTCGCTTT 196
QY 60 AlaArgArgTyGlySerTyGlySerGluGluGluLeuLeuLeuLeuLeuLeuLeuLeu 79
Db 197 GCTATCAGGCATCGGAAAGGATGATGACTCCGTGTAAGAGATCAACAAAGGTTGAGATA 256
QY 80 PheValGluAsnLeuAlaPheIleArgSerThrAsnArgGlyAspLeuSerTyThrLeu 99
Db 257 TTTTGGACAACTGAGATGATCCGATCGATACAGAAAGGACTATCATCAAACTC 316
QY 100 GlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAla 119
Db 317 GGTATCAATGAGTTTACCAGCTTACATGGATGATTCGTTAAACACAAAGTTGGGGCA 376
QY 120 AlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuPro 139
Db 377 TCTCAAACTGTCTCTGCCACCAAGGAATCTCAGCTCACTAACTGCTTCTGCA 436
QY 140 ValThrArgAspTrpArgGluGlnGlyIleValSerProValIleAspGlnGlySerCys 159
Db 437 GAGACGAAAGACTGGAGGAAGATGATTTAGCCAGTGAAGGACACAGGCAAGTGC 496
QY 160 GlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaTyThrGlnLeuThr 178
Db 497 GGATCTGTGGACATTCAGCACTCTGTGTGACTAGAGGAGCCCTATCCCAAGCATTT 556
QY 179 GlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsn 196
Db 557 GCGAAGGAATCTCTGTGACAGCAGCAGCTGTGTGACTGTGTGAGCTTTTAATAAC 616
QY 197 PheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyValIleTyAsnGlyCly 215
Db 617 TTTGGTTGCAATGAGGGTGGCTTCTCAAGCATTTTGAATATTAATTCATGTTGGT 676
QY 216 IleAspThrGluGlnThrTyProTyLeuGlyValMetGlyIleCysAsnPheIleGln 235
Db 677 CTGTGATTTGGTGTGACATATCATACCGGCAAGATGGCATATGTAATTCACAA 736
QY 236 GluAsnValGlyValIleAspSerIleAsnIleThrLeuGlyAlaGluAspGlu 255
Db 737 GCAATATTGCTGCTAAAGTCATCAGTTCTGTCAATATTACCTGGTGTCTGAATGAA 796
QY 256 LeuIleHisAlaValGlyLeuValArgProValSerValAlaPheGluValIleGly 275
Db 797 CTGAATACCGATTTGATTTGGTGTGCTGTTGCTTTTGGTGTGTAAGGG 856
QY 276 PheAsnLeuTyTrpIleGlyValTySerSerAspThrCysGlyArgAspProMetAsp 295
Db 857 TTCAACAGTACAGAGCGAGTTTACGCCAGCAGTGAATGTGGCAGACATCCCATGGAC 916
QY 296 ValAsnHisAlaValLeuAlaValGlyTyGlyValGluAspGlyIleProTyTrpLeu 315
Db 917 GTAAACCATGCTGTTCTGTGTGGTTACGGTGTGAAATGTTACTCCCTACTGGCTC 976
QY 316 IleIleAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyTrpPheIleMetGluLeuGly 335
Db 977 ATAAAGAACTCATGGGAGCAGATTTGGGTGAGATGATCTCAAAATGGAGATGGGA 1036
QY 336 LysAsnMetCysGlyValAlaThrCysAlaSerTyTrpIleValala 351
Db 1037 AAGAAATATGTGTGTGTGAACCTTGTGCATCTTACCCAACTGTTGCC 1084
RESULT 12
ADB94877
ID ADB94877 standard; cDNA; 1652 BP.

XX ADB94877;
AC 04-DEC-2003 (first entry)
XX Programmed cell death pathway cysteine protease cDNA #42.
XX
XX programmed cell death; plant development; plant cell cycle; ATL2; DAD1;
KW Dnase; lls; lsd1; nucellin-like aspartic protease; annexin; prohibitin;
KW fen-like protein; rac2; retinoblastoma-related protein; SINA; TFID;
KW TEGT; xylogenic Rnase; pur-alpha; cyteine protease; RPPs-like protein;
KW SP 91 NADPH oxidase subunit; NPS-like protein; BAG-1;
KW defender against cell death; lethal leaf spot; lesion stimulating death;
KW seven in absentia; transcription initiation factor;
KW testis enhanced gene transcript; Gene; ss.
XX Pinus radiata.
OS
XX US2003082724-A1.
FN
XX 01-MAY-2003.
PD
XX 14-AUG-2002; 2002US-00219220.
PF
XX 04-JUN-1999; 99US-00325932.
PR
XX (GENE-) GENESIS RES & DEV CORP LTD.
PA
XX Flinn B, Lasham A;
FI
XX WPI; 2003-786916/74.
DR P-PSDB; ADB94919.
XX
XX New isolated polynucleotide useful for modulating programmed cell death,
PT altering the development cycle of plant cells, and subsequently modifying
PT plant development.
XX
XX Claim 1; Page 154; 214pp; English.
XX
XX The invention describes an isolated polynucleotide (I) comprising a
CC nucleotide sequence that is one of 145 fully defined sequences of 221-
CC 3415 base pairs (bp), given in the specification, and/or its complements,
CC reverse complements, reverse sequences, or sequences having 75, 90 or 95
CC % sequence identity to or that hybridise under stringent hybridisation
CC conditions to one of the 145 sequences. The methods and compositions of
CC the present invention to do with (I) are useful for modulating programmed
CC cell death and thereby altering the development cycle of plant cells, and
CC altering plant development. This sequence encodes a protein associated
CC with the programmed cell death pathway.
XX
SQ Sequence 1652 BP; 476 A; 318 C; 393 G; 465 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1,81e-118 Length: 1652
Score: 1148.00 Matches: 228
Percent Similarity: 75.21% Conservative: 42
Best Local Similarity: 63.51% Mismatches: 79
Query Match: 61.55% Indels: 10
DB: 9 Gaps: 8

US-10-087-714-2 (1-352) x ADB94877 (1-1652)

QY 1 MetAlaAlaLysLeuLeuPhePheLeu-----LeuPheLeuValSerAlaLeuSerVal 18
Db 264 ATGGCTCGCGTCCTCGCCATTTGTTCTGAGCAGCTTTGTTTTCGTG---GCCGTATCTGTA 320
QY 19 AlaLeuAlaGly---PheGluGluAspAsnProIleArgSerValThrGlnArgProAsp 37
Db 321 TCTGCTGCGGGTCTGTTTGAAGAACTGGGTATATTGATATGTTGTCCTGACACAGATCAA 380
QY 38 SerIleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
Db 381 GACCTGGAGTCCAAATTCGTTAAATCTTGGAAACCAATCCCAAGTCTCTACAATTTGCG 440

PR	28-JUN-1999;	99US-0140823P.	PR	10-SEP-1999;	99US-0153070P.
PR	29-JUN-1999;	99US-0140991P.	PR	13-SEP-1999;	99US-0153758P.
PR	30-JUN-1999;	99US-0141287P.	PR	16-SEP-1999;	99US-0154018P.
PR	01-JUL-1999;	99US-0141842P.	PR	19-SEP-1999;	99US-0154039P.
PR	01-JUL-1999;	99US-0142154P.	PR	20-SEP-1999;	99US-0154779P.
PR	02-JUL-1999;	99US-0142055P.	PR	22-SEP-1999;	99US-0155139P.
PR	06-JUL-1999;	99US-0143290P.	PR	23-SEP-1999;	99US-0155486P.
PR	08-JUL-1999;	99US-0142803P.	PR	24-SEP-1999;	99US-0155659P.
PR	09-JUL-1999;	99US-0143920P.	PR	28-SEP-1999;	99US-0156458P.
PR	12-JUL-1999;	99US-0143977P.	PR	29-SEP-1999;	99US-0156596P.
PR	13-JUL-1999;	99US-0143542P.	PR	04-OCT-1999;	99US-0157117P.
PR	14-JUL-1999;	99US-0143624P.	PR	05-OCT-1999;	99US-0157753P.
PR	15-JUL-1999;	99US-0144005P.	PR	06-OCT-1999;	99US-0157865P.
PR	16-JUL-1999;	99US-0144085P.	PR	07-OCT-1999;	99US-0158029P.
PR	18-JUL-1999;	99US-0144086P.	PR	08-OCT-1999;	99US-0158232P.
PR	19-JUL-1999;	99US-0144325P.	PR	12-OCT-1999;	99US-0158369P.
PR	19-JUL-1999;	99US-0144331P.	PR	13-OCT-1999;	99US-0159293P.
PR	19-JUL-1999;	99US-0144332P.	PR	13-OCT-1999;	99US-0159294P.
PR	19-JUL-1999;	99US-0144333P.	PR	13-OCT-1999;	99US-0159295P.
PR	19-JUL-1999;	99US-0144334P.	PR	14-OCT-1999;	99US-0159329P.
PR	19-JUL-1999;	99US-0144335P.	PR	14-OCT-1999;	99US-0159330P.
PR	20-JUL-1999;	99US-0144352P.	PR	14-OCT-1999;	99US-0159331P.
PR	20-JUL-1999;	99US-0144632P.	PR	14-OCT-1999;	99US-0159637P.
PR	20-JUL-1999;	99US-0144884P.	PR	14-OCT-1999;	99US-0159638P.
PR	21-JUL-1999;	99US-0144814P.	PR	18-OCT-1999;	99US-0159584P.
PR	21-JUL-1999;	99US-0145086P.	PR	21-OCT-1999;	99US-0160741P.
PR	21-JUL-1999;	99US-0145088P.	PR	21-OCT-1999;	99US-0160767P.
PR	21-JUL-1999;	99US-0145085P.	PR	21-OCT-1999;	99US-0160768P.
PR	22-JUL-1999;	99US-0145087P.	PR	21-OCT-1999;	99US-0160770P.
PR	22-JUL-1999;	99US-0145089P.	PR	21-OCT-1999;	99US-0160814P.
PR	22-JUL-1999;	99US-0145192P.	PR	21-OCT-1999;	99US-0160815P.
PR	22-JUL-1999;	99US-0145145P.	PR	21-OCT-1999;	99US-0160815P.
PR	23-JUL-1999;	99US-0145218P.	PR	22-OCT-1999;	99US-0160980P.
PR	23-JUL-1999;	99US-0145224P.	PR	22-OCT-1999;	99US-0160981P.
PR	26-JUL-1999;	99US-0145276P.	PR	22-OCT-1999;	99US-0160989P.
PR	27-JUL-1999;	99US-0145313P.	PR	25-OCT-1999;	99US-0161404P.
PR	27-JUL-1999;	99US-0145918P.	PR	25-OCT-1999;	99US-0161405P.
PR	27-JUL-1999;	99US-0145919P.	PR	25-OCT-1999;	99US-0161406P.
PR	28-JUL-1999;	99US-0145951P.	PR	26-OCT-1999;	99US-0161359P.
PR	02-AUG-1999;	99US-0146386P.	PR	26-OCT-1999;	99US-0161360P.
PR	02-AUG-1999;	99US-0146388P.	PR	26-OCT-1999;	99US-0161361P.
PR	02-AUG-1999;	99US-0146389P.	PR	26-OCT-1999;	99US-0161361P.
PR	03-AUG-1999;	99US-0147038P.	PR	28-OCT-1999;	99US-0161920P.
PR	04-AUG-1999;	99US-0147204P.	PR	28-OCT-1999;	99US-0161920P.
PR	04-AUG-1999;	99US-0147302P.	PR	28-OCT-1999;	99US-0161993P.
PR	05-AUG-1999;	99US-0147192P.	PR	29-OCT-1999;	99US-0162142P.
PR	05-AUG-1999;	99US-0147260P.			
PR	06-AUG-1999;	99US-0147303P.			
PR	06-AUG-1999;	99US-0147416P.			
PR	09-AUG-1999;	99US-0147432P.			
PR	09-AUG-1999;	99US-0147935P.			
PR	10-AUG-1999;	99US-0148171P.			
PR	11-AUG-1999;	99US-0148319P.			
PR	12-AUG-1999;	99US-0148341P.			
PR	13-AUG-1999;	99US-0148565P.			
PR	13-AUG-1999;	99US-0148684P.			
PR	16-AUG-1999;	99US-0149368P.			
PR	17-AUG-1999;	99US-0149175P.			
PR	18-AUG-1999;	99US-0149426P.			
PR	20-AUG-1999;	99US-0149722P.			
PR	20-AUG-1999;	99US-0149939P.			
PR	20-AUG-1999;	99US-0149902P.			
PR	23-AUG-1999;	99US-0149930P.			
PR	25-AUG-1999;	99US-0150566P.			
PR	26-AUG-1999;	99US-0150884P.			
PR	27-AUG-1999;	99US-0151065P.			
PR	27-AUG-1999;	99US-0151080P.			
PR	30-AUG-1999;	99US-0151303P.			
PR	31-AUG-1999;	99US-0151438P.			
PR	01-SEP-1999;	99US-0151930P.			
PR	07-SEP-1999;	99US-0152363P.			
			Alignment Scores:		
			Pred. No.:		8.22e-114
			Score:		1108.00
			Percent Similarity:		45.22%
			Best Local Similarity:		39.06%
			Query Match:		59.41%
			DB:		3
			Length:		2149
			Matches:		241
			Conservative:		38
			Mismatches:		72
			Indels:		267
			Gaps:		6
			US-10-087-714-2 (1-352) x AAC44178 (1-2149)		
QY	1	MetAlaAlaLysLeuLeuPhe-----PheLeuLeuPheLeuValSerAlaLeuSer	17		
Db	81	ATGTCGCGAAACAATCCTATCATAGTAGTTTGGTGTCTCGTCGCCGCATCAGCA	140		
QY	18	ValAlaLeuAlaGlyPheGluGluAspAsnProIleArgSerValThrGlnArgProAsp	37		
Db	141	GCGGCTAATATCGGATTCGATGATCAACCCGATCCGAATGGTCTCCGATGGTCTCCGG	200		
QY	38	SerIleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla	57		
Db	201	GAGGTAGAAGATCTGTTTCCAGATCTTAGTCAATCTCGTCACGTTCTCTCTTCGCT	260		
QY	58	ArgPheAlaArgMetTyrGlySerTyrGlySerGluGluIleLeuValLysArgPhe	77		
Db	261	CGCTTCACTACCGATATGGTAAAGATATCAGAACGTGGAGGAGATGAAGCTTCGATTC	320		
QY	78	GlyIlePheValGluAsnLeu-----	84		

Db	321	TCGATTTTCAAGGAGATCTTGATGGTAATGACTCCAACTTATGATAGTGTATTTATGTT	380
Qy	84	-----	84
Db	381	CAGATAATGCCGATGACTTTGTCTATGCTGCTCCACCGATTTTGAGAACACAGCGACTT	440
Qy	84	-----	84
Db	441	CCGTCCACAGCGTCCAGGTGCTGCTCTCAGATTTCAGGTTATGCGGCTCAATTCGCTGCGT	500
Qy	84	-----	84
Db	501	ATATCGCTTGCTGATTAGTGTCAGCTTTCCCTTCAGCGGGATTTCATACAGCGCCAGCC	560
Qy	84	-----	84
Db	561	ATCGCTCATCATATCACCAGTCAAAAGGTGACAGCAGGCTCATAGACGCCCGCAGCT	620
Qy	85	-----	86
Db	621	CGCCATAGTGGTTTACCGAATACGTGCGCAACACCGTCTTCGGAGAGCTGTCTATACGC	680
Qy	86	-----	86
Db	681	GTAACACAGCGAGCGCTGGCGCGATTTAGCCCGACATAGCCCGACTGTTTCGTCCTATTC	740
Qy	86	-----	86
Db	741	CGCGCAGCAGTACGTCAGTCCCGGCTGTATGCGCGAGGTACCGACTGCGGCTGAG	800
Qy	86	-----	86
Db	801	TTTTTTAAGTACGTAATAATCGTTGTGAGGCCAACGCCATAATGCGGGCTGTTGCGCGG	860
Qy	86	-----	86
Db	861	CATCCAAAGCCATTCATGGCCATATCAATGATTTCTGGTGGTACCGGGTTGAGAACGG	920
Qy	86	-----	86
Db	921	GTGTAAGTGAATGCACTTGCCATGTTTTCAGCGCAGTGAGCAGAGATAGCGCTGATGT	980
Qy	86	-----	86
Db	981	CCGCGGTGCTTTTGGCTTACGCACACCCCGTCACTAGCTGAACAGGAGGACAGCTG	1040
Qy	86	-----	86
Db	1041	ATAGAAACAGAGCCACTGGAGCACCTCAAAACACCATCATACATAATCAGTAAGTT	1100
Qy	87	-----	98
Db	1101	GGCAGCATCCAAATCTTGATTTGATCAGATCCCAACAGAAAGCTTATCTTACAA	1160
Qy	99	LeuGlyIleAenGlnPheAlaAspLeuThrTTPGluPheArgThrAsnArgLeuGly	118
Db	1161	CTCGGTGTTAATCAATTTGCTGATTGACATGGCAGAGTTTCAAGGACCAAGCTTGGT	1220
Qy	119	AlaAlaGlnAenCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeu	138
Db	1221	GCTGCTCAGAACTGCTCTGCCACTTTAAAGGGCAGCCACCAAGGTCTACAGAAAGCTCTT	1280
Qy	139	ProValThrArgAspTTPArgGluGlnGlyIleValSerProValIleAspGlnGlySer	158
Db	1281	CCTGAACAAAGACTGGAGAGAGATTGTATGTTAGTCCGTCGCAAGATCAGGAGGT	1340
Qy	159	CysGlySer---TTPThrPheSerThrThrGlyAlaLeuGluAlaAlaThrGlnLeu	177
Db	1341	TGTGGATCTTGTGACATTCAGCAGCTGAGCTCTTTCAGGAGCTTACCATCAGGCA	1400
Qy	178	ThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsn	195
Db	1401	TTTGGAAAGAAATATCTCTCTGTAGCAACAGCTTGTGGATGTGCTGGAGCTTTCAAT	1460

RESULT 14

AAC44535

ID AAC44535 standard; DNA; 1072 BP.

XX

AC AAC44535;

XX

DT 18-OCT-2000 (first entry)

XX

DE Zea mays DNA fragment SEQ ID NO: 43187.

XX

KW Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway; metabolic; pathway;

KW promoter; termination sequence; corn; ss.

XX

OS Zea mays subsp. mays.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-00301439.

XX

PR 25-FEB-1999; 99US-0121825P.

PR

PR 05-MAR-1999; 99US-0123180P.

PR

PR 09-MAR-1999; 99US-0123548P.

PR

PR 23-MAR-1999; 99US-0125788P.

PR

PR 25-MAR-1999; 99US-0126264P.

PR

PR 29-MAR-1999; 99US-0126785P.

PR

PR 01-APR-1999; 99US-0127462P.

PR

PR 08-APR-1999; 99US-0128234P.

PR

PR 16-APR-1999; 99US-0129845P.

PR

PR 19-APR-1999; 99US-0130077P.

PR

PR 21-APR-1999; 99US-0130449P.

PR

PR 23-APR-1999; 99US-0130510P.

PR

PR 28-APR-1999; 99US-0131443P.

PR

PR 30-APR-1999; 99US-0132048P.

PR

PR 04-MAY-1999; 99US-0132484P.

PR 05-MAY-1999; 99US-0132485P. 99US-0145087P.
PR 06-MAY-1999; 99US-0132486P. 99US-0145089P.
PR 06-MAY-1999; 99US-0132487P. 99US-0145192P.
PR 07-MAY-1999; 99US-0132863P. 99US-0145145P.
PR 11-MAY-1999; 99US-0134256P. 99US-0145218P.
PR 14-MAY-1999; 99US-0134218P. 99US-0145224P.
PR 14-MAY-1999; 99US-0134219P. 99US-0145276P.
PR 14-MAY-1999; 99US-0134221P. 99US-0145913P.
PR 14-MAY-1999; 99US-0134370P. 99US-0145918P.
PR 18-MAY-1999; 99US-0134768P. 99US-0145919P.
PR 19-MAY-1999; 99US-0134941P. 99US-0145951P.
PR 20-MAY-1999; 99US-0135124P. 99US-0146386P.
PR 21-MAY-1999; 99US-0135353P. 99US-0146388P.
PR 24-MAY-1999; 99US-0135629P. 99US-0146389P.
PR 25-MAY-1999; 99US-0136021P. 99US-0147038P.
PR 27-MAY-1999; 99US-0136392P. 99US-0147204P.
PR 28-MAY-1999; 99US-0136782P. 99US-0147302P.
PR 01-JUN-1999; 99US-0137222P. 99US-0147192P.
PR 03-JUN-1999; 99US-0137528P. 99US-0147260P.
PR 04-JUN-1999; 99US-0137502P. 99US-0147303P.
PR 07-JUN-1999; 99US-0137724P. 99US-0147416P.
PR 08-JUN-1999; 99US-0138094P. 99US-0147493P.
PR 10-JUN-1999; 99US-0138540P. 99US-0147935P.
PR 10-JUN-1999; 99US-0138847P. 99US-0148171P.
PR 14-JUN-1999; 99US-0139119P. 99US-0148319P.
PR 16-JUN-1999; 99US-0139452P. 99US-0148341P.
PR 18-JUN-1999; 99US-0139453P. 99US-0148565P.
PR 17-JUN-1999; 99US-0139492P. 99US-0148684P.
PR 18-JUN-1999; 99US-0139454P. 99US-0149368P.
PR 18-JUN-1999; 99US-0139460P. 99US-0149175P.
PR 18-JUN-1999; 99US-0139461P. 99US-0149930P.
PR 18-JUN-1999; 99US-0139462P. 99US-0150566P.
PR 18-JUN-1999; 99US-0139463P. 99US-0150884P.
PR 18-JUN-1999; 99US-0139750P. 99US-0151065P.
PR 18-JUN-1999; 99US-0139457P. 99US-0151066P.
PR 18-JUN-1999; 99US-0139458P. 99US-0151066P.
PR 18-JUN-1999; 99US-0139459P. 99US-0151080P.
PR 18-JUN-1999; 99US-0139460P. 99US-0151030P.
PR 22-JUN-1999; 99US-0139899P. 99US-0151303P.
PR 23-JUN-1999; 99US-0140353P. 99US-0151438P.
PR 23-JUN-1999; 99US-0140354P. 99US-0151930P.
PR 24-JUN-1999; 99US-0140695P. 99US-0152363P.
PR 28-JUN-1999; 99US-0140823P. 99US-0153070P.
PR 29-JUN-1999; 99US-0140921P. 99US-0153758P.
PR 30-JUN-1999; 99US-0141287P. 99US-0154018P.
PR 01-JUL-1999; 99US-0141842P. 99US-0154039P.
PR 01-JUL-1999; 99US-0142154P. 99US-0154779P.
PR 02-JUL-1999; 99US-0142055P. 99US-0155139P.
PR 06-JUL-1999; 99US-0142390P. 99US-0155486P.
PR 08-JUL-1999; 99US-0142803P. 99US-0155659P.
PR 09-JUL-1999; 99US-0142920P. 99US-0156458P.
PR 12-JUL-1999; 99US-0142977P. 99US-0156596P.
PR 13-JUL-1999; 99US-0143542P. 99US-0157117P.
PR 14-JUL-1999; 99US-0143624P. 99US-0157753P.
PR 15-JUL-1999; 99US-0144005P. 99US-0157865P.
PR 16-JUL-1999; 99US-0144085P. 99US-0158029P.
PR 16-JUL-1999; 99US-0144086P. 99US-0158232P.
PR 19-JUL-1999; 99US-0144325P. 99US-0158369P.
PR 19-JUL-1999; 99US-0144331P. 99US-0159293P.
PR 19-JUL-1999; 99US-0144332P. 99US-0159294P.
PR 19-JUL-1999; 99US-0144333P. 99US-0159295P.
PR 19-JUL-1999; 99US-0144334P. 99US-0159330P.
PR 19-JUL-1999; 99US-0144335P. 99US-0159331P.
PR 20-JUL-1999; 99US-0144352P. 99US-0159637P.
PR 20-JUL-1999; 99US-0144632P. 99US-0159638P.
PR 20-JUL-1999; 99US-0144884P. 99US-0159584P.
PR 21-JUL-1999; 99US-0144814P. 99US-0160741P.
PR 21-JUL-1999; 99US-0145086P. 99US-0160767P.
PR 21-JUL-1999; 99US-0145088P. 99US-0160768P.
PR 22-JUL-1999; 99US-0145085P. 99US-0160768P.
PR 22-JUL-1999; 99US-0145087P. 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P. 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P. 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P. 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P. 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P. 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P. 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P. 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P. 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P. 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P. 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P. 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P. 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P. 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P. 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P. 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P. 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P. 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P. 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P. 99US-0147935P.
PR 09-AUG-1999; 99US-0148171P. 99US-0148319P.
PR 11-AUG-1999; 99US-0148319P. 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P. 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P. 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P. 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P. 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P. 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P. 99US-0149426P.
PR 20-AUG-1999; 99US-0149723P. 99US-0149723P.
PR 20-AUG-1999; 99US-0149723P. 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P. 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P. 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P. 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P. 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P. 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P. 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P. 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P. 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P. 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P. 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P. 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P. 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P. 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P. 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P. 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P. 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P. 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P. 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P. 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P. 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P. 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P. 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P. 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P. 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P. 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P. 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P. 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P. 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P. 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P. 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P. 99US-0159295P.
PR 14-OCT-1999; 99US-0159330P. 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P. 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P. 99US-0159637P.
PR 18-OCT-1999; 99US-0159638P. 99US-0159638P.
PR 21-OCT-1999; 99US-0160741P. 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P. 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P. 99US-0160768P.

D6 661 AGAAGACTCATGGCGGCTGACTGGGGTGATGAGGGTTACTTCAAGATGGMAATGGGCACA 720

QY 337 AsnMetCysGlyValAlaThrCysAlaSerTyProIleValala 351
:::|||||
D6 721 AACATGTGCGGTGTGCTACGTGTGCATCCTACCCCTGTTGTGCA 765

RESULT 15
ACL19838/c
ID ACL19838 standard; DNA; 618 BP.
XX
AC ACL19838;
DT 17-OCT-2003 (first entry)
XX
DE DNA clone originating in barley containing SNP encoding sequence #9829.
XX
KW Barley; single nucleotide polymorphism; SNP; genotype-phenotype analysis;
KW gene; ss.
OS Hordeum vulgare.
XX
PN W02003057877-A1.
XX
PD 17-JUL-2003.
XX
PF 16-DEC-2002; 2002WO-IB005403.
XX
PR 20-DEC-2001; 2001JP-00387059.
PR 20-DEC-2001; 2001JP-00387131.
PR 20-DEC-2001; 2001JP-00403299.
PR 20-DEC-2001; 2001JP-00403300.
PR 27-SEP-2002; 2002JP-00327515.
XX
PA (UYNI-) UNIV JAPAN OKAYAMA.
XX
PI Sato K, Takeda K, Kohara Y;
DR WPI; 2003-587127/55.
XX
PT Single nucleotide polymorphism sites in barley varieties and DNA
PT sequences containing them for analysis and identification of barley
PT varieties and production of barley transformants with desired
PT characteristics.
XX
PS Disclosure; SEQ ID XX; 284pp; Japanese.
XX
CC The present invention relates to oligonucleotide clones originating in
CC barley (Hordeum vulgare) which contain single nucleotide polymorphisms
CC (SNP). The oligonucleotides may be used for analysis of SNPs among barley
CC varieties, identification of particular varieties and genotype-phenotype
CC analysis, isolation of specific genes and creation of new varieties by
CC transformation of barley varieties with them and production of new barley
CC varieties with desired properties. The present sequence represents an
CC oligonucleotide clone DNA sequence featured in the specification. The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published-pct-sequences
XX
SQ Sequence 618 BP; 142 A; 178 C; 152 G; 145 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.: 3,48e-85 Length: 618
Score: 846.00 Matches: 154
Percent Similarity: 87.13% Conservative: 22
Best Local Similarity: 76.24% Mismatches: 22
Query Match: 45.36% Indels: 4
DB: 8 Gaps: 3

US-10-087-714-2 (1-352) x ACL19838 (1-618)

Qy 154 LysAspGlnGlySerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAla 172
:|||||
Db 616 AAAAACCCGCCCATTCGGCTCTGCTGGACGTTCAGCTACTCTGGCAGCTTAGGCA 172

QY 173 AlaTyrThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCys 190
Db 556 GCATATACTCAGGCCACTGGAAGAACATCTCTCTTCAGCAACAGCTGGTTGACTGT 497
QY 191 AlaSerAlaPheAsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyr 209
Db 496 GTGTGGTGGGTTCAATAACTTTGGATGCAACGAGAGCCTTCCATCCAGGGGTTTGAGTAC 437
QY 210 ValLysTyrAsnGlyGlyIleAspThrGluGlnThrTyrProTyrLeuGlyValMetGly 229
Db 436 ATCAATACAAATGCGCGGATCGACACCGAGAGTCTCTACCTTACAAGGGTGTCAATGGT 377
QY 230 IleCysAsnPhelYsGlnGlnAsnValGlyValLysValIleAspSerIleAsnIleThr 249
Db 376 GTCGTGCATTACAGCTGAAATGCTGCGGTTTCAGGTTTGGACTCAGTAAACATCACA 317
QY 250 LeuGlyAlaGluAspGluLeuLysHisAlaValGlyLeuValArgProValSerValAla 269
Db 316 CTGAATGCCGAGGATGAAGACTGAAGAACGCGGTTGGGTTGCGCCAGTGAGTGTGGCC 257
QY 270 PheGluValValLysGlyPheAsnLeuTyrLysGlyValTyrSerSerAspThrCys 289
Db 256 TTTCAGGTGATTACCGGTTTCAGGCAGTACAGAGCGGAGTTTACACGAGCGCCATTGT 197
QY 290 GlyArgAspProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAsp 309
Db 196 GGCACCTACCCCTCATGACGTGAACCAACGCGGCTTGGCAGTCGGCTATGCTGTGGAGAAC 137
QY 310 GlyIleProTyrTrpLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyr 329
Db 136 GCGGTCCCGTATTGGCTCATCAAGAACTCGTGGGGCGCCGACTGGGGTGACAAATGGTTAC 77
QY 330 PheLysMetGluLeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIle 349
Db 76 TTCAGATGGAATGGCAAGAACATGTGGCTATTGCTACCTGGCGCTCTTACCCTGTC 17
QY 350 ValAla 351
Db 16 GTGGCG 11

Search completed: March 29, 2004, 22:35:16
Job time : 383 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 29, 2004, 22:22:45 ; Search time 77 Seconds
(without alignments)

2536.918 Million cell updates/sec

Title: US-10-087-714-2

Perfect score: 1865
Sequence: 1 MAAKLFFLLFLYSALVAL.....ELGKMGVATCASYPIVAV 352

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 692709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10087714/runat_26032004_113453_967/app_query.fasta_1.519
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10087714@cgn1_1_44@runat_26032004_113453_967 -NCFU=6 -ICFU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1300.5	69.7	1407	4	US-09-325-932A-107
2	656.5	35.2	677	4	US-09-325-932A-111
3	594	31.8	1600	4	US-09-325-932A-98
4	590	31.6	1463	4	US-09-325-932A-92
5	583	31.3	1390	3	US-08-821-994-61
6	583	31.3	1441	3	US-08-821-994-63
7	581	31.2	1434	3	US-08-821-994-62
8	580	31.1	1428	4	US-09-325-932A-118
9	573.5	30.8	1577	3	US-08-821-994-59
10	567	30.4	1056	1	US-09-500-651-1
11	567	30.4	1056	1	US-08-813-591-1
12	562	30.1	1474	3	US-08-821-994-64

13	557.5	29.9	1661	3	US-08-821-994-82	Sequence 82, Appl
14	557	29.9	1102	3	US-08-821-994-86	Sequence 86, Appl
15	557	29.9	1553	3	US-08-821-994-60	Sequence 60, Appl
16	549.5	29.5	1366	3	US-08-883-526-2	Sequence 2, Appl
17	549.5	29.5	1366	4	US-09-519-283-2	Sequence 2, Appl
18	547.5	29.4	1839	3	US-09-120-365-4	Sequence 4, Appl
19	547.5	29.4	1839	3	US-09-515-039-4	Sequence 4, Appl
20	542	29.1	1833	3	US-09-120-365-6	Sequence 6, Appl
21	542	29.1	1833	3	US-09-515-039-6	Sequence 6, Appl
22	535.5	28.7	1915	3	US-09-120-365-2	Sequence 2, Appl
23	535.5	28.7	1915	3	US-09-515-039-2	Sequence 2, Appl
24	535.5	28.7	1916	3	US-09-120-365-88	Sequence 88, Appl
25	535.5	28.7	1916	3	US-09-515-039-88	Sequence 88, Appl
26	530	28.4	2167	4	US-09-325-932A-102	Sequence 102, App
27	527.5	28.3	990	3	US-08-964-308-12	Sequence 12, Appl
28	527.5	28.3	990	3	US-08-964-308-13	Sequence 13, Appl
29	527.5	28.3	990	3	US-08-964-313-12	Sequence 12, Appl
30	527.5	28.3	990	3	US-08-964-313-13	Sequence 13, Appl
31	527.5	28.3	990	4	US-09-069-138-12	Sequence 12, Appl
32	527.5	28.3	990	4	US-09-069-138-13	Sequence 13, Appl
33	527.5	28.3	1482	1	US-08-330-121B-1	Sequence 1, Appl
34	527.5	28.3	1482	4	US-08-536-861-1	Sequence 1, Appl
35	527.5	28.3	1482	5	PCT-US95-13820-1	Sequence 1, Appl
36	527.5	28.3	1619	1	US-08-208-007A-1	Sequence 1, Appl
37	527.5	28.3	1619	4	US-08-915-095A-1	Sequence 1, Appl
38	527.5	28.3	1619	4	US-08-798-096-1	Sequence 1, Appl
39	527.5	28.3	1619	4	US-08-798-095A-1	Sequence 1, Appl
40	527.5	28.3	1619	4	US-09-953-956-1	Sequence 1, Appl
41	527.5	28.3	1619	4	US-08-553-125A-1	Sequence 1, Appl
42	527.5	28.3	1619	4	US-10-114-464-1	Sequence 1, Appl
43	527.5	28.3	1619	5	PCT-US94-04781-1	Sequence 1, Appl
44	527.5	28.3	1669	3	US-08-964-308-3	Sequence 3, Appl
45	527.5	28.3	1669	3	US-08-964-313-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-09-325-932A-107

; Sequence 107, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: death and their use in the modification of forestry plant devel
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325.932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 107
; LENGTH: 1407
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-09-325-932A-107

Alignment Scores:			
Pred. No.:	9,25e-148	Length:	1407
Score:	1300.50	Matches:	247
Percent Similarity:	79.7%	Conservative:	33
Best Local Similarity:	70.37%	Mismatches:	60
Query Match:	69.73%	Indels:	11
DB:	4	Gaps:	5

US-10-087-714-2 (1-352) x US-09-325-932A-107 (1-1407)

QY	9	LeuteuPheLeuValSerAlaLeuSerValAlaLeuAlaGlyPheGluAspAsnPro 28
DB	91	CTCCCTCCGCGCGCGTGTCTCCGCCGCGTGTCTCGAGAGTCCACCCC 150
QY	29	lleArgSerValThrGlnArgProAsp-----SerIleGluProAlaIleLeu 44

```

/genes="At5g60360"
73. .1149
/genes="At5g60360"
/codon_start=1
/evidence=experimental
/product="putative cysteine proteinase AAP"
/protein_id="AAN31819.1"
/db_xref="GI:23397068"
/translations="MSAKTILSGVILVLVAASAAANIGFDSNPIMVSDGLREVVEE
SVSQILQGRHVLSPARFTHRYGKYQVNEEMKLRPSIFKENLDLIRSTNKKGLSKYL
GVNQFADLTQWERQTKLGAQNCSTLKGSHKVTPEALPETKDMREDGIVSPVKDQG
GCSCSWFTFTTGALEAYHQARKGISLSEQLVDCAGAFNNYGCNGGILPSQAFYIK
SNGLDTERKAFPTIGKDETCSTSAENVGVVLNSVNTITLCAEDELKHAVLGVRPVSIA
FEVHFRIKYSVYTDHSCGSTFMDNVHNAVLAGVGYVEDGVFVYMLIKNSWGADWDGDK
GYFKMEMGNMCMGIATCASYPVVA"
misc_difference 1128
/genes="At5g60360"
/notes="compared to genomic sequence"
/replace="a"
1150. .1409
/genes="At5g60360"
misc_difference 1393
/genes="At5g60360"
/notes="not present in genomic sequence"

ORIGIN
Alignment Scores:
Pred. No.: 6,29e-107 Length: 1409
Score: 1279.50 Matches: 240
Percent Similarity: 77.93% Conservative: 39
Best Local Similarity: 67.04% Mismatches: 72
Query Match: 68.61% Indels: 7
DB: 8 Gaps: 4

US-10-087-714-2 (1-352) x BT000673 (1-1409)
Qy 1 MetAlaAlaLysLeuLeuPhe-----PheLeuLeuPheLeuValSerAlaLeuSer 17
Db 73 ATGCTCGGAAACAAATCCATCATCAGTAGTTTGGTGGTCTCTCGTCGCCGATCAGCA 132
Qy 18 ValAlaLeuAlaGlyPheGluGluAspAsnProIleArgSerValThrGlnArgProAsp 37
Db 133 CGCGCTAATATCGGATTCGATGAGTCAACACCCGATCCGAATGGTCTCCGATGCTCCCG 192
Qy 38 SerIleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
Db 193 GAGTAGAAGATCTGTTCCAGATCTTAGGTCATCTCGTCACGTTCTCTCTTTGCT 252
Qy 58 ArgPheAlaArgArgTyrGlyLysSerTyrGlySerGluGluIleLysLysArgPhe 77
Db 253 CGCTTCACTCACCGATATGTTAAAGATATCAGAACGTCGAGGAGATGAAGCTTCGATTC 312
Qy 78 GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyr 97
Db 313 TCGATTTTCAAGGAGATCTTGATTTGATCAGATCCACCAACCAAGAGGCTTATCTTAC 372
Qy 98 ThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeu 117
Db 373 AAACCTCGGTGTAAATCAATTTGCTGATTTGACATGGCAAGAGTTTCAAGACCAAGCTT 432
Qy 118 GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyVal 137
Db 433 GGTGCTGCTCAGAACTGCTCTGCCACTTTTAAAGGGCAGCCCAAGGTCACAGAGCAGCT 492
Qy 138 LeuProValThrArgAspTrpArgGluGlnGlyIleValSerProValLysAspGlnGly 157
Db 493 CTTCTGTAACAAAAGACTGGAGAGAGATGGTATCGTTAGTCCGGTCAAAAGATCAGGGA 552
Qy 158 SerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGln 176
Db 553 GGTGTGGATCTTGTGACATTCAGCACAACTGAGGCTCTTGAGGCGAGCTTACCATCAG 612
Qy 177 LeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPhe 194

```

Search completed: March 29, 2004, 23:30:32

Job time : 3309 secs

6

```

Db 613 GCATTTGAAAAGAAATTTCTCTCTGAGCAACAGCTTGTGGATTGTGTGGAGCTTTC 672
Qy 195 AsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsn 213
Db 673 AATTAATATGTTGCAATGTTGGCTTCTCTCAAGCCTTGTATATACATCAATCCAAC 732
Qy 214 GlyGlyIleAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyIleCysAsnPhe 233
Db 733 GGTGGCTCGACACAGAGAAAGCTTATCCTTATACCGGTAAAGATGAACCTGCAAAATTT 792
Qy 234 LysGlnGluAsnValGlyValLysValIleAspSerIleAsnIleThrLeuGlyValGlu 253
Db 793 TCAGCTGAAAACGTTGGTGTACAAGTCTCACTCACTCAACATTAATCTCTGGGTGCTGAA 852
Qy 254 AspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGluValVal 273
Db 853 GATGAACCTGAAGCATGCGGTTGGATTGTCACGGCCAGTAAAGCATAGCATTTGAGGTTATA 912
Qy 274 LysGlyPheAsnLeuTyrLysGlyValTyrSerSerAspThrCysGlyArgAspPro 293
Db 913 CACTCGTTCGGCTTTACAGAGTGGAGTTTACACTGATAGTCACTGTGGNAGTACTCCA 972
Qy 294 MetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIleProTyr 313
Db 973 ATGGATGTGAACACCGCGTTTTCGCGTTTATGGAGTTGAAGACCGGTGTACCATAT 1032
Qy 314 TrpLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLysMetGlu 333
Db 1033 TGGCTTATTAAGAACTCATGGGGAGCGGATTTGGGGCCGACAAAGGTTACTTCAAGATGAG 1092
Qy 334 LeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIleValAla 351
Db 1093 ATGGGGAAGAACATGTGTGGTATTGCTACATGTGGTGCATACCCCGTTGTGGCT 1146

```

Db 151 ATCCGGCTCTTC-----CCGACGGCGGCTCCGCGACCTCGAGTCTCCATCCTC 201
Qy 45 GlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAlaArgTyrGly 64
Db 202 CAGATCGTGGCGCGCACCGCCGACGCTTCTCTTCGCGCGCTTCGCCAACAGGTATGG 261
Qy 65 LysSerTyrGlySerGluGluGluLeuLysLysArgPheGlyLeuPheValGluAsnLeu 84
Db 262 AAGAGGTACGAGACCGCGGAGGATCAAGCTCGGTTCCAGATCTTCAGGGAGATCTC 321
Qy 85 AlaPheIleArgSerThrAsnArgLysAspLeuSerTyrThrLeuGlyLeuAsnGlnPhe 104
Db 322 AGTTGATCCGATCCACCAACAGAGGCTTCGCCCTACACCTCGGTGTCATTAAGTT 381
Qy 105 AlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAlaAlaGlnAsnCysSer 124
Db 382 GCTGATGGAGCTGGGAGGATTCAGAGGCACAGACTGGAGCTGCTCAAAAGTCTCT 441
Qy 125 AlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProValThrArgAspTrp 144
Db 442 GCCACCAACCAAGGCGCAACACAGCTCACCGACGAGCTCTCCCGAGATGAAGACTGG 501
Qy 145 ArgGluGlnGlyLeuValSerProValLysAspGlnGlySerCysGlySer---TrpThr 163
Db 502 AGAAGAAAGGCGATGTAAGCCCAATTAAGATCAGGSCACTGGGATCTTGTGGACT 561
Qy 164 PheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGlySer-----Thr 181
Db 562 TTCAGTACCACCTGGAGCTCTTGAGGCTGCTTATACCAAGCATTCGGGAAACAAATCTCT 621
Qy 182 LeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPheGlyCys---Gly 200
Db 622 CTGCTGAGCAGCAGCTGTGGACTGTGTGGGCTTTCAACAATTTGGATGTAGTGGT 681
Qy 201 GlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGlyLeuAspThrGluGln 220
Db 682 GGAGTGCATCCCAAGCTTTGAGTACGTCAAGTACACCGTGGCTTCATACCGAGGAA 741
Qy 221 ThrTyrProTyrLeuGlyValMetGlyIleCysAsnPheLysGlnGluAsnValGlyVal 240
Db 742 GCATTCCTTATACCGCAGTGATGTAGCTGCMAAATTCGCGCTGATTAATGTGTGTCT 801
Qy 241 LysValIleAspSerIleAsnIleThrLeuGlyAlaGluAspGluLeuLysHisAlaVal 260
Db 802 CAAGTGTCTGACACTGTGTAACATCACCTTGGGTGCTGAGATGAACATAAGCATGCAGTT 861
Qy 261 GlyLeuValArgProValSerValAlaPheGluValValLysGlyPheAsnLeuTyrLys 280
Db 862 GCCTTCGTCGGCCAGTGTGTGGCATTCAGGTCTGTGAAGACTTCAGATGTGTACAG 921
Qy 281 LysGlyValTyrSerSerAspThrCysGlyArgAspProMetAspValAsnHisAlaVal 300
Db 922 TCGGGTGTCTACACGAGCGATACATCGGTAGCATTCCATCGATGTGAACCATGCTGT 981
Qy 301 LeuAlaValGlyTyrGlyValGluAspGlyLeuProTyrTrpLeuIleLysAsnSerTrp 320
Db 982 CTCGCTGTGTTATGGAGTGAAGATGGTGTTCGCTTCGCTCATCAAGAATTCCTGG 1041
Qy 321 GlyThrAsnTrpGlyAspAsnGlyTyrPheLysMetGluLeuGlyLysAsnMetCysGly 340
Db 1042 GAGCAGACTGGGTGACCGATACCTACTTCAAGATGAGATGGGAAGAACATGTGTGA 1101
Qy 341 ValAlaThrCysAlaSerTyrProIleValAla 351
Db 1102 GTGCTACTTGTGCATCATACCTCTGTGTGGCC 1134

RESULT 2
US-09-325-932A-111
; Sequence 111, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette

; TITLE OF INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: death and their use in the modification of forestry plant deve:
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 111
; LENGTH: 677
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-09-325-932A-111
Alignment Scores:
Pred. No.: 3,86e-70 Length: 677
Score: 656.50 Matches: 131
Percent Similarity: 71.68% Conservative: 31
Best Local Similarity: 57.96% Mismatches: 49
Query Match: 35.20% Indels: 15
DB: 4 Gaps: 4
US-10-087-714-2 (1-352) x US-09-325-932A-111 (1-677)
Qy 3 AlalysLeuLeuPhePheLeuLeuPheLeuValSerAlaLeuSerValAlaLeuAlaGly 22
Db 26 GCTAGATTCTGTGTTCT 85
Qy 23 PheGluAspAsnProIleArgSerValThrGlnArgProAspSerIleGluProAla 42
Db 86 TTCAGGCGCGCGAC-----CTCGAGTCTCTCC 112
Qy 43 IleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAlaArg 62
Db 113 ATCTCCAAACCGTGTGCCAGCGCGCTCCGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 172
Qy 63 TyrGlyLysSerTyrGlySerGluGluIleLysLysArgPheGlyIlePheValGlu 82
Db 173 TACGAGAAGAGGTACGAGACGCGAGAGATCAAGTTGAGGTTGCGATAATTACAGGGAG 232
Qy 83 AsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyrThrLeuGlyIleAsn 102
Db 233 AATCTCAAGCTCATTCGATCCACCAAGAGGGCTTGGCTTACACTCTCGCTGTAT 292
Qy 103 GlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAlaAlaGlnAsn 122
Db 293 CAGTATGCTGACTGGAGCTGGAGGAGTTCAAGACGCGACAGACTGGGAGCTTCTCAAGAC 352
Qy 123 CysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProValThrArg 142
Db 353 TGCTCTCCACCCCAAGGCGCGACCAAGCTCACCGACGCTGTCTTCTCCAAACAGAA 412
Qy 143 AspTrpArgGluGlnGlyIleValSerProValLysAspGlnGlySerCysGlySer--- 161
Db 413 GACTGGAGAAAGAGGCGCATTTAAGCCCATTTAAATCAAGCGGCTGTGGATCTTGC 472
Qy 162 TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGlySer--- 180
Db 473 TGGAGTTTCAAGCGCCTCGAGCTCTCGAGGCTGCTTATCACCAGACACCGGAAAGGA 532
Qy 181 ---ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPheGlyCys 199
Db 533 ATCTCTCTGTCTGAGCAGCAGCTCGTGAATCGGCTACGCGCTTCAACAATTTGGATGC 592
Qy 200 ---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyIleAspThr 218
Db 593 GATGCGGGTTGCGGTCACAAAGCCTTCGAGTACATCAAGTACACCGTGGCTTGGAGACC 652
Qy 219 GluGlnThrTyrProTyr 224
Db 653 GAGGAAGCTTATCTTAT 670
RESULT 3
US-09-325-932A-98

```

; Sequence 98, Application US/09325932A
; Parent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Laeham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: death and their use in the modification of forestry plant develop
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 98
; LENGTH: 1600
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
; US-09-325-932A-98

Alignment Scores:
Pred. No.: 5,77e-62 Length: 1600
Score: 594.00 Matches: 147
Percent Similarity: 55.28% Conservative: 57
Best Local Similarity: 39.84% Mismatches: 121
Query Match: 31.85% Indels: 44
DB: 4 Gaps: 15

US-10-087-714-2 (1-352) x US-09-325-932A-98 (1-1600)

QY 5 LeuLeuPhePheLeuLeuPheLeuValSerAlaLeuSerValAlaLeuAlaGlyPheGlu 24
Db 116 CTGCGGTGCCCTGACGACGACGCGCTCCGCCCTCGACATGTCCATCGTCAGTACGAT 175
QY 25 GluAspAsnProIleArgSerValThrGln-----ArgProAspSerIleGlu 40
Db 176 CGGGCCCAACGGCAGCGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 235
QY 41 ProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAla 60
Db 236 ATGGCGCGTC-----TACGAGAGCTGGCTC 259
QY 61 ArgArgTyrGlyIySerTyrGlySerGluGluGluIleIySerIySerGlyPheGlyPhe 80
Db 260 GCCAAGCAGCGCAAGCGCTTCAACGCCCTGGCGGAGAAAGAGAGCGCTTCCAGGTCCTC 319
QY 81 ValGluAsnLeuAlaPheIleArgSerThrAsnArg---LysAspLeuSerTyrThrLeu 99
Db 320 AAGGACAACCTCCGTTTCATCGACGACCAACAGCGCGGCGGACCGGACCTACACGGTC 379
QY 100 GlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAla 119
Db 380 GGCGCTCAACAGATTCCCGACCTCATTAACGAGAGGTACCGGTCCATGTACCTGGCGGCC 439
QY 120 AlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArg-----Phe 133
Db 440 AGGATGATCGTTCGGCGCGCGCGCTCGGAGGGCCCGACGATCGTACGCGGTGGCC 499
QY 134 ValAspGlyValLeuProValThrArgAspTyrArgGluGlnGlyIleValSerProVal 153
Db 500 GCCCGGGAGAGCTCGCGCGCTCCGTCGATGGAGGAAGGAGCGCGCTTGTTCACGTC 559
QY 154 LysAspGlnGlySerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAla 172
Db 560 AAGGACCAAGGAGCTCGGAGAGTTCTGGCGCTTCTTACATTCCTCTCTGTGGAGGG 619
QY 173 AlaTyrThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCys 190
Db 620 ATAAACAAGCTTGTGATGTGATTTGATCTCTCTCGAGCAGGAACTTGTGGACTGC 679
QY 191 AlaSerAlaPheAsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyr 209
Db 680 GATACATCTTACAAATGAA---GGATGAATGGCGGCTCATGGATTATCCCTTTGAATTC 736
QY 210 ValIyetyrAsnGlyGlyIleAspThrGluGlnThrTyrProTyrLeuGlyValMetGly 229

```

```

Db      737 ATTATCAACACGAGGACATTGATACCGAGAGGAGATTATCCCTATAGAGCTGTAGATAGC 796
Qy      230 ILeCysAsnPhelYsGlnGluAsnValGlyVallYsValIleAspSerIle---AsnIle 248
Db      797 ACTTGTGACCAATACAGAGNAGACGCAAGGTTGTGACGATTGACGATTATCAAGATGTT 856
Qy      249 ThrLeuGlyAlaGluAspGluLeuLysHisAlaValGlyLeuValArgProValSerVal 268
Db      857 CCAGAAATGATGAGAAGCAATCCAAAGGCGAGTTGCT---AATCAACCACTCAGTGTG 913
Qy      269 AlaPheGluVal---ValLysGlyPheAsnLeuTyrLysGlyValTyrSerSerAsp 287
Db      914 GCCATTGAAGCAGGAGCGGGAATTCAGTTTTATGATCGGTATATTTACTGGC--- 970
Qy      288 ThrCysGlyArgAspProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyVal 307
Db      971 AAATGTGGG-----ACAGCTCTGGATCATGGGTTACTGCAGTCGGATATGGCACA 1021
Qy      308 GluAspGlyIleProTyrTrpLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsn 327
Db      1022 GAAACCGGAGTTGATTACTGGATAGTGAAGAACTCATGGGCGGTAGCTGGGAGAGCAA 1081
Qy      328 GlyTyrPheLysMet-----GluLeuGlyLysAsnMetCysGly 340
Db      1082 GGGTACATCAAAATGGCAGCAAGATGTGGCCAATAGCCCCACTGGCAAA-----TGTGGT 1135
Qy      341 ValAlaThrCysAlaSerTyrProIle 349
Db      1136 ATAGCAATGGAGGCTCTCTACCCCATC 1162

RESULT 4
US-09-325-932A-92
; Sequence 92, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: death and their use in the modification of forestry plant devel
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 92
; LENGTH: 1463
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-09-325-932A-92

Alignment Scores:
Pred. No.: 1,53e-61 Length: 1463
Score: 590.00 Matches: 155
Percent Similarity: 57.45% Conservative: 57
Best Local Similarity: 42.01% Mismatches: 120
Query Match: 31.64% Indels: 37
DB: 4 Gaps: 16

US-10-087-714-2 (1-352) x US-09-325-932A-92 (1-1463)

```

[illegible]

```

QY 61 rg-ArgTyrGlyLysSerTyrGlySerGluGluGluLeuLysLysArgPheGlyIlePhe 80
Db 183 GCCAGCAGCGAAGCCCTACAAACGCGCTTCGCGAGCGGAGCGCGGTTCGAGATCTTC 242
QY 81 ValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyrThrLeuGly 100
Db 243 CGCGACAACTCCGTTCTCGACGAGCAACAGCCCTCAACCGGTCTGACACCTTCGG 302
QY 101 IleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAla 119
Db 303 ATGACCGGTTCGCCGACCTCACCAACGAGGAGTACCGCGCGTCTTACCTCGGCACCGG 362
QY 120 -----AlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheVal 134
Db 363 AGCGACCCATCGCGCGGTGGGAGAGCGCGCGCGAGCGCGCTACCGCCCGCGC 422
QY 135 ---AspGlyValLeuProValThrArgAspTrpArgGluGlnGlyIleValSerProVal 153
Db 423 CCGGACGACATGCTCGCGGAGCGCGTCACTCGGAGCAACCGCGCTTAAACAGGTC 482
QY 154 LysAspGlnGlySerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAla 172
Db 483 AAGGACCGAGGAGCTTGGGAGAGCTGCTGGGCGCTTCTTACCATAGCTGCTGGAGGA 542
QY 173 AlaTyrThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCys 190
Db 543 ATCAACAGATCGTACCGCGGAGTTCATATCCCTCCGAGCAAGAGCTCGTGGATTGT 602
QY 191 AlaSerAlaPheAsnAsnGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyr 209
Db 603 GACCGGCGCTTAC---GATCGCGGTTCGCAATGGTGGGCTCATGGACTATCGCTTCCAGTTC 659
QY 210 ValLysTyrAsnGlyIlePheThrGluGlnThrTyrProTyrLeuGlyValMetGly 229
Db 660 ATCATTCGATGGTGGCATGTGACATGATGAGGACTACTCTTACACGGGAGTGGATGGA 719
QY 230 IleCysAsnPheLysGlnGluAsnValGlyValLysValIleAspSerIle---AsnIle 248
Db 720 ACCTGTGATGCGTCTTAAGTTAACTCAAAGGTGGTGGAGCACTTATGGGTACGAGGATGC 779
QY 249 ThrLeuGlyAlaGluAspGluLeuLysHisAlaValGlyLeuValArgProValSerVal 268
Db 780 CCGCGCTTTGACGAGAGAGCGCTTGAAGAGCGTGTGGCT---CATCAACCGGTGATGTT 836
QY 269 AlaPheGluVal---ValLysGlyPheAsnLeuTyrLysLysGlyValTyrSerSerAsp 287
Db 837 GCCATTGAAGCTGGAGGAGAGATTTCCAACTTATGAAACCGCGGTGTTCACTGGAGAA 896
QY 288 ThrCysGlyArgAspProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyVal 307
Db 897 ---TGTGGA-----ACTGCATGGACACGCGTGTGATCGCAGTTGGATACGGGAGG 944
QY 308 GluAspGlyIleProTyrTrpLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsn 327
Db 945 CAACATGGTGTGCTACTGCTTGTGAAGAACTCGTGGGCTCCTTGTGGGCGGAGAGC 1004
QY 328 GlyTyrPheLysMetGlu-----LeuGlyLysAsn-----MetCysGlyValAla 342
Db 1005 GGATACATCAAGATGGAGAGAACTTGGCCCAACAACACTACTTTGGCAAGTGGGTATCGCA 1064
QY 343 ThrCysAlaSerTyrProIle 349
Db 1065 ATGGAGGCTTCTTACCTGTTC 1085

```

RESULT 5

```

US-08-821-994-61
; Sequence 61, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier R
; APPLICANT: Jenson, Ian
; TITLE OF INVENTION: Promoters

```

```

; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 61
; LENGTH: 1390
; TYPE: DNA
; ORGANISM: Brassica napus
US-08-821-994-61

```

Alignment Scores:

```

Pred. No.: 9,97e-61 Length: 1390
Score: 583.00 Matches: 142
Percent Similarity: 56.65% Conservative: 71
Best Local Similarity: 37.77% Mismatches: 117
Query Match: 31.26% Indels: 46
DB: Gaps: 17

```

US-10-087-714-2 (1-352) x US-08-821-994-61 (1-1390)

```

QY 2 AlaAlaLysLeuLeuPhePheLeuLeuValSerAlaLeuSerValAlaLeuAla 21
Db 51 TCAGCGAAACTCCTCTCTTACTTCTCTTTGTCGTC-----TTCAATTTCATTAGCC 101
QY 22 GlyPheGluGlu-----AspAsnProIleArgSerValThrGlnArgProAsp 37
Db 102 TCCAGTAGTAGTCCATCATCAACGACCAACATCTCTTCTCCATCTGACCGTCTGG 161
QY 38 SerIleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
Db 162 AGAACCGATGAAGAAGTGTATGTCATC-----TACTTA 194
QY 58 ArgPheAlaArgArgTyrGlyLysSerTyrGlySerGluGluGluIle----- 73
Db 195 AATGGTCTTGGAGCAGCGGAAAGTAACAGCAACAGCAACGATTTATCAACCAACAA 254
QY 74 LysLysArgPheGlyIlePheValGluAsnLeuAlaPheIle-----ArgSerThrAsn 91
Db 255 GATGAAGATCAATATTTTCAAGACAACTAAGATTTCATCGATCTACACAGAGAGAAC 314
QY 92 ArgLysAspLeuSerTyrThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGlu 111
Db 315 AACAAAGACGCTGATCTTACAAAGCTTGGTCTAACCATATTTCGCTGATCTCCTAACGATGAG 374
QY 112 PheArgThrAsnArgLeuGlyAla-----AlaGlnAsnCysSerAlaThrAla 127
Db 375 TACCGAGTTTATACCTCGGCGGAGAACCGAGCGCTGTCCGCGCATCACTAAGGCCAAG 434
QY 128 HisGlyAsnHisArgPheValAspGlyVal-----LeuProValThrArgAsp 143
Db 435 AACGTTAACTGAATACTACTCAGCGCGAGTAACAGCACTGGAGGTTCCGAGACCGTTGAC 494
QY 144 TrpArgGluGlnGlyIleValSerProValLysAspGlnGlySerCysGlySer---Trp 162
Db 495 TGGAGACAGAGAGGAGCGGTTAATCCCATTAACCAAGAGATCTTTCGGAAGTGTGG 554
QY 163 ThrPheSerThrThrGlyAlaLeuGluAlaTyrThrGlnLeuThrGlySer----- 180
Db 555 GCGTTTTCACAGCTGCAGCAGTAGAGGCATATAACAAAGATCGTAACAGGAGAGCTCAT 614
QY 181 ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPheGlyCys--- 199
Db 615 TCTCTGTCCGAACAAGAACTTGTGCACTGCGCAAAATCATACCAACAA---GGCTGTAAAC 671
QY 200 GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGlyIleAspThrGlu 219
Db 672 GCGCGTCTAATGATTATGCTTTTCAATTTCATCATGAAACCGCGGATTAACACCGAG 731

```



```
/ TITLE OF INVENTION: Promoters
/ FILE REFERENCE: PPD 50108
/ CURRENT APPLICATION NUMBER: US/08/821,994A
/ CURRENT FILING DATE: 1997-03-22
/ EARLIER APPLICATION NUMBER: PCT/GB97/00729
/ EARLIER FILING DATE: 1997-03-18
/ EARLIER APPLICATION NUMBER: GB 9606062.9
/ EARLIER FILING DATE: 1996-03-22
/ NUMBER OF SEQ ID NOS: 89
/ SOFTWARE: Patent in Ver. 2.1
/ SEQ ID NO 62
/ LENGTH: 1434
/ TYPE: DNA
/ ORGANISM: Brassica napus
/ US-08-821-994-62

Alignment Scores:
Pred. No.: 1434
Score: 581.00
Percent Similarity: 56.68%
Best Local Similarity: 38.24%
Query Match: 31.15%
DB: 3

US-10-087-714-2 (1-352) x US-08-821-994-62 (1-1434)

QY 4 LysLeuLeuPhePheLeuValSerAlaLeuSerValAlaLeuAala---Gly 22
Db 57 AAATCTCTCTCTTACTTCTCTTATACGTC-----TTCTTCTGTTAGCTCCGGT 107
QY 23 PheGlu-----GluAspAsnProIleArgSerValThrGlnArgProAspSerile 39
Db 108 TATGAGTCCATCATCAGTGAACACCTCTCAGTCTCCATCTCCATCAGCGTTCTGGTGAAC 167
QY 40 GluProAlaLeuLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPhe 59
Db 168 GATGAAGAAGTATATCCATC-----TACTTAAGATGG 200
QY 60 AlaArgArgTyrglylysserTyrglyserGluGluGluile-----LysLys 75
Db 201 TCCTTGAGCAGCGGAAAGTAACAGCAACAGCAACGATATTATCAACCAACAGCAAGAA 260
QY 76 ArgPheGlylePheValGluAsnLeuAlaPhele-----ArgSerThrAsnArgLys 93
Db 261 AGATTCAATATTTCAAAGACAACTAAGATTCTATCATGATCATACACACAGAGAAACAAG 320
QY 94 AspLeuSerTyrglyleAsnGlnPheAlaAspLeuThrTyrglyleGluPheArg 113
Db 321 AACGCTACTTACAGCTTGTCTTACCATATTTGCTGATCTCTACATGATGATGATGATG 380
QY 114 ThrAsnArgLeuGlyAla-----AlaGlnAsnCysSerAlaThrAlaHisGly 129
Db 381 AGTTTATACCTCGGGCAAGAACCGAGCTGTCTCGCGCATCTACTAAGGCCAAGAACGTT 440
QY 130 AsnHisArgPheValAspGlyVal-----LeuProValThrArgAspTyrArg 145
Db 441 AACATGAATACTCAGCGCGAGTAACAGCTGGAGGTTCCGGAGCGTGTGCTGAGAGA 500
QY 146 GluGlnGlyIleValSerProValIleAspGlnGlySerCysGlySer---TrpThrPhe 164
Db 501 AAGAAAGGAGCGGTTAATGCCATTAAAGACCAAGAACTTCCGGAAGTGTGTTGGGGGTTT 560
QY 165 SerThrThrGlyAlaLeuGluAlaAlaTyrglyleThrGlnLeuThrGlySer-----ThrLeu 182
Db 561 TCAACAGCTGAGCAGTATAGAGGATATAACAAAGATCGTAAACAGAGAACCTCGTATCTTTG 620
QY 183 SerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPheGlyCys---GlyGly 201
Db 621 TCCGACACAGAACTTGTGCTGCTGCGACAAATCGTACAAACCAA---GGCTGTAAACGGGGT 677
QY 202 LeuProSerGlnAlaPheGluTyrglyleAsnGlyIleAspThrGluGlnThr 221
Db 678 CTAATGGATTATGCTTTTCAATTATGATAAAGACCGCGGATTAATAACACCGAGAAAGAC 737
```

```
QY 222 TyrProTyrglyleGlyValMetGlyIleCysAsnPhelysGlnGluAsnValGlyValLys 241
Db 738 TATCTTTACCACGGAACCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATG 797
QY 242 ValIleAspSerile---AsnIleThrLeuGlyAlaGluAspGluLeuLysHisAlaVal 260
Db 798 ACTATCGATGATACGAGATGTTCTTCTAGTAAAGATGAAACCGGTTTGAAGAGCAGTT 857
QY 261 GlyLeuValArgProValSerValAlaPheGluVal---ValLysGlyPheAsnLeuTyrg 279
Db 858 TCA---TACCAGCGCTGTGAGTGTGCTTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 914
QY 280 LysLysGlyValTyrglyleAspThrCysGlyArgAspProMetAspValAsnHisAla 299
Db 915 CAATCTGGAATCTTCACTGGA---AAGTGTGGT-----ACGAATATGGATCAGCT 962
QY 300 ValLeuAlaValGlyTyrglyleGluAspGlyIleProTyrglyleLeuLysAsnSer 319
Db 963 GTGGTGGCGGTGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1022
QY 320 TrpGlyThrAsnTrpGlyAspAsnGlyTyrglylePheLysMetGluLeuGly----- 335
Db 1023 TGGGTGATACGTTGGGAGAGGATGTTATAGATGAGATGAGAGAGAAACGTTGGCTCTAAA 1082
QY 336 LysAsnMetCysGlyValAlaThrCysAlaSerTyrglyle 349
Db 1083 TCCGTAAGTGTGGATTCGATAGAGCCTCGTATCCGGTT 1124

RESULT 8
US-09-325-932A-118
/ Sequence 118, Application US/09325932A
/ Patent No. 6451604
/ GENERAL INFORMATION:
/ APPLICANT: Flinn, Barry
/ APPLICANT: Lasham, Annette
/ TITLE OF INVENTION: Compositions affecting programmed cell
/ TITLE OF INVENTION: death and their use in the modification of forestry plant deve
/ FILE REFERENCE: 1022
/ CURRENT APPLICATION NUMBER: US/09/325,932A
/ CURRENT FILING DATE: 1999-06-04
/ NUMBER OF SEQ ID NOS: 206
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 118
/ LENGTH: 1428
/ TYPE: DNA
/ ORGANISM: Pinus radiata
/ US-09-325-932A-118

Alignment Scores:
Pred. No.: 2,416-60
Score: 580.00
Percent Similarity: 53.41%
Best Local Similarity: 39.51%
Query Match: 31.10%
DB: 4

US-10-087-714-2 (1-352) x US-09-325-932A-118 (1-1428)

QY 2 AlaAlaLysLeuLeuPhePheLeuLeuValSerAlaLeuSerValAlaLeuAala 21
Db 19 TCCTCCAGCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 78
QY 22 GlyPheGluGluAspAsnProIleArgSerValThrGlnArgProAspSerileGluPro 41
Db 79 ATAAAGCGCTGAGCATGACCTTATTCGTAAGTACGAGGAGAGTAAATCA---GATCCA 135
QY 42 AlaIleLeuGlyValLeuGlySerCysArgHisAla---PheHisPheAlaArgPheAla 60
Db 136 CAGATCTCTGAGTCTCGGAGCGCCCTGTTCAACGCTGAAGCGCATTTGAGGGCTTTATC 195
QY 61 ArgArgTyrglylysserTyrglyserGluGluGluLeuLysLysArgPheGlylePhe 80
```

196 AGGCGCTACGGGAAGAAGTACTCGGGGCGGAGACGACGACGACCGCTTCGGTGTCTTC 255
Qy 81 ValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyThrLeuGly 100
Db 256 AAGAGCAATTTACTAAGAGCCTTGGAGCACCAGAGGCTCGACCCCGAGGCTCCCATGGC 315
Qy 101 IleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAlaAala 120
Db 316 GTCACAGAAATCTCTGATTTGACACAAGAGGAGTTCGACGACAGTAGTATCTAGGCGCTCAGG 375
Qy 121 GlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProVal 140
Db 376 GCACACCGATCCGAGAGCGCCACGATGCTCCAAATTTTGCCCAAAACGATCTGCCGGAG 435
Qy 141 ThrArgAspTrpArgGluGlnGlyLeuValSerProValLysAspGlnGlySerCysGly 160
Db 436 GAGTTCGATTCGAGAGAGAGGAGCGCTGACCGAGGTTAAGATCAGGGATCGTGGGT 495
Qy 161 Ser---TrpThrPheSerThrThrGlyAlaLeuGluAlaLalaTyThrGlnLeuThrGly 179
Db 496 TCCTGCTGGGCTTTTCAGCACAAACCGGGCGGTAGAGGGCGCGAATTTTCCTGAAGACGGG 555
Qy 180 Ser-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsn----- 195
Db 556 AAGCTGGTGACCTTGAGGAGCAACAATTTGGTGGACTGGGATCAGAGTGGGATCCTTCG 615
Qy 196 -----AsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTy 209
Db 616 GACGCAAGATCATGTGATCTGTTGCAATGCTGGCTAATGCTGGCTAATGCTGCTTATCAATAT 675
Qy 210 ValLysTyAsnGlyGlyIleAspThrGluGlnThrTyProTyThrLeuGlyValMetGly 229
Db 676 GCTCTGAAGCTGGTGGATTTGGAGAGAGAGGAGTACCCATATATCTGGAAAGACGGA 735
Qy 230 IleCysAsnPheLysGlnGluAsnValGlyValLysValIleAspSerIleAsnIleThr 249
Db 736 ACTTGACGCTTTAACAAGAACAAAATTTGTCACAGGTT-----TCGAATTTTCAGC 786
Qy 250 LeuGlyAlaGluAspGluLeuLysHisAlaValGlyLeuValArg-----ProValSer 267
Db 787 GTTGTCTTATTGATGAGATCAATTTGCTGCAAAATCTGGTGAAGATGGACCTCTATCA 846
Qy 268 ValAlaPheGluValValLysGlyPheAsnLeuTyLysLysGlyValTySerSerAsp 287
Db 847 GTGGGAATCAATGTGCA---TTTATGCAGACATACGTAGGAGGTGTATCTTSCCCATAC 903
Qy 288 ThrCysGlyArgAspProMetAspValAsnHisAlaValLeuAlaValGlyTyGly--- 306
Db 904 ATCTGCAGCAAGCGCATGTTGGAT-----CATGGTGTGCTCTCTGGTGGATATGGTTCT 957
Qy 307 -----ValGluAspGlyIleProTyThrTrpLeuIleLysAsnSer 319
Db 958 CGAGGCTTGTCTCCATAGTAAGAGCAAA---CCCTACTGGATCATAAAGACTCA 1014
Qy 320 TrpGlyThrAsnTrpGlyAspAsnGlyTyThrPheLysMetGluLeuGlyLysAsnMetCys 339
Db 1015 TGGGACCTAACTGGGAGAAATGGATTCTACAAACTTTGACGGGGACATAAGTTTGC 1074
Qy 340 GlyValAlaThrCysAlaSer 346
Db 1075 GGAATCAACAACATGGTTTCC 1095

RESULT 9

US-08-821-994-59
; Sequence 59, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A

; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 59
; LENGTH: 1577
; TYPE: DNA
; ORGANISM: Brassica napus
US-08-821-994-59

Alignment Scores:

Pred. No.: 1,73e-59 Length: 1577
Score: 573.50 Matches: 141
Percent Similarity: 55.28% Conservative: 58
Best Local Similarity: 38.17% Mismatches: 114
Query Match: 30.75% Indels: 47
DB: 3 Gaps: 13

US-10-087-714-2 (1-352) x US-08-821-994-59 (1-1577)

Qy 5 LeuLeuPhePheLeuLeuPheLeuValSerAlaLeuSerValAlaLeuAlaGlyPheGlu 24
Db 63 CTCACCTTCTTCCCTTCTTCTAGTTCTTCTCTGAGCTTCTCATCATCATCTTCCGAT 122
Qy 25 GluAspAsnProIleArgSerValThrGlnArgProAspSerIleGluProAlaIleLeu 44
Db 123 GACATCTCCGAGCTG----- 137
Qy 45 GlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAlaArgArgTyGly 64
Db 138 -----TTTCAGCGTTCGTCAGAGACACGCGC 164
Qy 65 LysSerTyGlySerGluGluGluIleLysLysArgPheGlyIlePheValGluAsnLeu 84
Db 165 AAAACGTCAGCTTCGGAGGAGAGAGACACACAGGATTCGAATCTTTAAAGACATCAC 224
Qy 85 AlaPheIle---ArgSerThrAsnArgLysAspLeuSerTyThrLeuGlyIleAsnGln 103
Db 225 GACTTGTGTACACGACACACAAACATCGCTAACTCTACTTCTCTCTCATCTCAATGCC 284
Qy 104 PheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAlaAlaGlnAsnCys 123
Db 285 TTCGGGATCTGACTCACCACGAGTTCAAGGCTCTGCTCTGGAGGATCTCTGCTTCT 344
Qy 124 SerAla-----ThrAlaHisGlyAsnHis---ArgPheValAspGlyValLeuPro 139
Db 345 TCAGCTCCTTTCGTGCTAAGGACAGAGCTTTGAGAACGTTTCGGGGAAGGTTCCA 404
Qy 140 ValThrArgAspTrpArgGluGlnGlyIleValSerProValLysAspGlnGlySerCys 159
Db 405 GATCTGTTGATGGAGGAGAAAGAGGCTGTACTAATGTCAAGATCAAGAAAGTGC 464
Qy 160 GlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaLalaTyThrGlnLeuThr 178
Db 465 GGAGCGTGTGCTTCTTCGCGGACTGGAGCTATGGAAGGAATCAACAGATTGTAAACA 524
Qy 179 GlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsn 196
Db 525 GGAGATCTCATCGCTCTCTGAGCAAGAACATCTGATTTGTAAGTCATACACCAT 584
Qy 197 PheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyThrValLysTyAsnGlyGly 215
Db 585 ---GGATGCAATGGTGGTCTCATGGACTACGCTTTTCAATTTGTCTATAAAACCATGG 641
Qy 216 IleAspThrGluGlnThrTyProTyThrLeuGlyValMetGlyIleCysAsnPheLysGln 235
Db 642 ATTGACACAGAGAAAGATTATCTTATCAAGAACGCTGATGGCACCTGTAGAAAGATAG 701
Qy 236 GluAsnValGlyValLysValIleAspSer---IleAsnIleThrLeuGlyAlaGluAsp 254


```
Db 702 TTGAATAGAACGGTGTGACAAATTGATAGCTACGCTGGTGTAAATCAATCAATGACGAGAA 761
Qy 255 GluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPhe---GluValVal 273
Db 762 GCGTTACTAGAACGCTGACG---GCTCAGCCAGTTAGTGTGGTATCTGTGGGAGCGAG 818
Qy 274 LysGlyPheAsnLeuTyrLysLysGlyValTyrSerSerAspThrCysGlyArgAspPro 293
Db 819 AGAGCGTTTCAGTTATCTTAAAGGATATCTCT-----GGCCCATGTTCA 866
Qy 294 MetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyLeuProTyr 313
Db 867 ACATCATTTGACCCAGCAGTCTCATCTGATGATACGTTCAAGAACGGTGTGATTAC 926
Qy 314 TrpLeuLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheIysMetGlu 333
Db 927 TCGATCGTGAAGAACTCTTGGGGAAGAGTTGGGAATGGATGGGTTTATCCACATGCGAG 986
Qy 334 Leu-----GlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProfile 349
Db 987 CGTAACACCGGCAACGACGAGAGTATCGGGAATCAACATGCTGCTTCATATCCCATC 1046

RESULT 10
US-09-500-651-1
; Sequence 1, Application US/09500651
; GENERAL INFORMATION:
; APPLICANT: ASANO, MINAO
; APPLICANT: KAWAI, MITSUKO
; APPLICANT: MIWA, TETSUYA
; APPLICANT: NIO, NORIKI
; TITLE OF INVENTION: AMINOPEPTIDASE GX, AND A METHOD OF
; TITLE OF INVENTION: HYDROLYSING A PROTEIN WITH THE SAME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESS: P.C. JEFFERSON DAVIS HIGHWAY, SUITE 400
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/500,651
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/813,591
; FILING DATE:
; APPLICATION DATA: JP 030458/1997
; FILING DATE: 14-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-845-0
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1056 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORGANISM: Glycine max
; FEATURE:
```

```
; NAME/KEY: CDS
; LOCATION: 1..1056
; OTHER INFORMATION: /note= "METHOD OF DETERMINING THE
; OTHER INFORMATION: CHARACTERISTICS: P"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..9
; OTHER INFORMATION: /note= "INSERTION SEQUENCE, METHOD
; OTHER INFORMATION: OF DETERMINING THE CHARACTERISTICS: E"
US-09-500-651-1
Alignment Scores:
Pred. No.: 5,63e-59 Length: 1056
Score: 567.00 Matches: 145
Percent Similarity: 56.65% Conservative: 51
Best Local Similarity: 41.91% Mismatches: 103
Query Match: 30.40% Indels: 47
DB: 16 Gaps: 16
US-10-087-714-2 (1-352) x US-09-500-651-1 (1-1056)
Qy 32 ValThrGlnArgProAspSerIleGluProAlaIleLeuGlyValLeuGlySerCysArg 51
Db 9 GTACGACAGCGCCA-----CGCGACAGGCGCGC 38
Qy 52 HisAla-----PheHisPheAlaArgPheAlaArg-ArgTyrG1 64
Db 39 CACGTTGGCCACCGAGGAGGAGCTGATGCCATGTACGAGCAGTGTCTGTGAAGCAGCG 98
Qy 64 YLysSerTyrGlySerGluGluGluLeuLysLysArgPheGlyLlePheValGluAsnLe 84
Db 99 GAAGGTGTACACGCGCTCGCGAGGAAGAGCGCTTCCAGATCTTCAAGACAACT 158
Qy 84 uAlaPheIleArgSerThrAsn---ArgLysAspLeuSerTyrThrLeuGlyIleAsnG1 103
Db 159 GCGATTTCATCGACGACCACTCCCGGGAGGACCGCACTACCACTCGGACTGAACCG 218
Qy 103 nPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAlaAla-----G1 121
Db 219 GTTCGCTGATCTCACCACGAGGAATACAGGCGCAAGTACTTGGGAACCAAGATCATCC 278
Qy 121 nAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyVal-----Le 138
Db 279 CAACCGGAGCTCGGAAGACCCCGGAGCAACCGCTACGCGCCGCGCGGACAAAT 338
Qy 138 uProValThrArgAspTrpArgGluGluGlnGlyIleValSerProValLysAspGlnGlySe 158
Db 339 GCCTGATTCGCTTCAATTGGAGGAAGAGGTGCTGTCTCTCTGTCAAAAGACCAAGGAG 398
Qy 158 rCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGlnLe 177
Db 399 CTGTGGGAGCTGTGGGCATCTCAGCAATCGGTGAGTAGAAGGATAAATAAGATAGT 458
Qy 177 uThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAs 195
Db 459 AACAGCGCAACTGATTTGCTTATCAGAACCAAGAAATGGTGGATGTGATACTGGATATAA 518
Qy 195 nAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnG1 214
Db 519 CCAA---GGATGCAATGGAGACTTATGGACTATGATTTTGAATTCATATCAACATGG 575
Qy 214 YGlyIleAspThrGluGlnThrTyrTrpTyrLeuGlyValMetGlyIleCysAsnPhely 234
Db 576 CGGCATTGATCTGATGAGGATTACCATACCGGTGTGTGTGATGTAGTAGCGACACATA 635
Qy 234 sGlnGluAsnValGlyValLysValIleAspSerIleAsnIleThrLeuGlyAlaGluAs 254
Db 636 TAGGAAAAATGCTAAAGTCTCTTCTATGTAGTACTACGAA---GATGTTCTCTCCCTATGA 692
Qy 254 pGlu-----LeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGlu-- 271
Db 693 TGAGTTAGCCTTGAAAAGGCCGCTTGA---AATCAGCCGCTGAGCGTGTCTATTGAAGG 749
```


Db 750 AGGGGGCAGGAATTCATTATATATCTGGTGTATTCAGG-----GGGAG 797
 Qy 291 GASPProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyI 311
 Db 798 ATGTGCACAGCAGCTAGATCATGTGCTGTGGTGTGGTATGGAACAGCTAAAGGTCA 857
 Qy 311 eProTyrTrpLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLys 331
 Db 858 TGATTATTGGATCGTAAGCAATTCATGGGTTCTAGCTGGGAGAGGATGGCTACATCAG 917
 Qy 331 sMetGluLeuGlyLysAsnMet-----CysGlyValAlaLysCys 344
 Db 918 ATTAGAA-----AGAAATCTTGCTAACAGCAGATCAGCAAGTGTGGAATTCGAATTGA 971
 Qy 344 salaserTyrProile 349
 Db 972 GCCATCTTATCCCTT 987

RESULT 12

US-08-821-994-64
 ; Sequence 64, Application US/08821994A
 ; Patent No. 6228643

GENERAL INFORMATION:

; APPLICANT: Greenland, Andrew J
 ; APPLICANT: Thomas, Didier RP
 ; APPLICANT: Jepson, Ian
 ; TITLE OF INVENTION: Promoters
 ; FILE REFERENCE: PPD 50108
 ; CURRENT APPLICATION NUMBER: US/08/821,994A
 ; CURRENT FILING DATE: 1997-03-22
 ; EARLIER APPLICATION NUMBER: PCT/GB97/00729
 ; EARLIER FILING DATE: 1997-03-18
 ; EARLIER APPLICATION NUMBER: GB 9606062.9
 ; EARLIER FILING DATE: 1996-03-22
 ; NUMBER OF SEQ ID NOS: 89
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 64

LENGTH: 1474

TYPE: DNA
 ORGANISM: Brassica napus

US-08-821-994-64

Alignment Scores:

Pred. No.: 3,86e-58 Length: 1474
 Score: 562.00 Matches: 144
 Percent Similarity: 55.41% Conservative: 66
 Best Local Similarity: 37.99% Mismatches: 117
 Query Match: 30.13% Indels: 52
 DB: 3 Gaps: 19

US-10-087-714-2 (1-352) x US-08-821-994-64 (1-1474)

Qy 2 AlaAlaLysLeuLeuPhePheLeuLeuPheLeuValSerAlaLeuSerValAlaLeuAla 21
 Db 42 TCAACAAACTCATTTCTTACTTCTCTATACGTC-----GTCGTTTCATTAGCC 92
 Qy 22 GlyPheGluGluAspAsnProileArgSerValThrGlnArgProAspSerileGluPro 41
 Db 93 TCCGGTGATGATGCCACCATCACTTAAACACCATCTCAATCTTCCATCCGAC----- 143
 Qy 42 AlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHis-----PheAla 57
 Db 144 -----GGCTCATGGAGAACCGATGACAGAACTGAGGTCCTACTTACTTA 185
 Qy 58 ArgPheAlaArgArgTyrGlyLys-----SerTyrGlySerGluGluLys 74
 Db 186 CAGTGGTGTGGGAGCAGCGGAAACTAGCAACACACAGGTATCGTCAACCAACAGAC 245
 Qy 75 LysArgPheGlyLePheValGluAsnLeuAlaPheIle-----ArgSerThrAsnArg 92
 Db 246 GAAAGTTCATATTTTCAAAGAACACCTAAGGTTTCAATTGATCATACCAATGAGAACAC 305

Qy 93 LysAspLeuSerTyrThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPhe 112
 Db 306 AAGAACCTACTTACAAAGCTTGGTCTCACCATTATCTCTGATCTCACTAAGCATGATGATC 365
 Qy 113 ArgThrAsnArgLeuGlyAla-----AlaGlnAsnCysSerAlaThrAlaHis 128
 Db 366 CGGAGGTATACCTTCGGGGGCAAGAACCGAGTCTGTCCGCCATCTACTAAGGCCNAGAAC 425
 Qy 129 GlyAsnHisArgPheValAspGlyVal-----LeuProValThrArgAspTrp 144
 Db 426 GTTAACATGAATACTTCGGCCGCGAGTAAACAGCAGCTGGAGGTTCCGGAGACGGTTGATTGG 485
 Qy 145 ArgGlnGlnGlyIleValSerProValLysAsnGlnGlySerCysGlySer---TrpThr 163
 Db 486 AGCGAAAGAGAGCCGTTAATGCCATTAAACCAAGAACTTCCGGAAGTTGTTGGCG 545
 Qy 164 PheSerThrThrGlyAlaLeuGluAlaLysTyrThrGlnLeuThrGlySer-----Thr 181
 Db 546 TTTTCGACAGCTGCAGCAGTAGAAGGTATAAACAGATCGTAACAGGAGAACTCATATCT 605
 Qy 182 LeuSerGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPheGlyCys---Gly 200
 Db 606 CTGTCGAAACAAGAACTTGTGCACTGGCAGCAGATCTCTACAAACAA---GGCTGCAACGGT 662
 Qy 201 GlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGlyIleAspThrGluGln 220
 Db 663 GGTTAATGGACTATGCTTTTCAATTCATCATGAAAAACGGCGTTTGAACACCGAGCAA 722
 Qy 221 ThrTyrProTyrLeuGlyValMetGlyIleCysAsnPhelysGlnGluAsnValGlyVal 240
 Db 723 GATTATCTTACCCTGCTGTTCCAAATGGAATGCAATCTTTCTGGAAGAATTCAGAGTT 782
 Qy 241 LysValIleAspSerIleAsnIleThrLeuGlyAlaGluAspGlu-----LeuLysHis 258
 Db 783 GTAACCTATTGATGTTACGAA---GATGTTCTTACTGAAGATGAACGGCTTCAAGAGA 839
 Qy 259 AlaValGlyLeuValArgProValSerValAlaPheGluVal---ValLysGlyPheAsn 277
 Db 840 GCAGTTTCA---TACCAGCCCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 896
 Qy 278 LeuTyrLysLysGlyValTyrSerSerAspThrCysGlyArgAspProMetAspValasn 297
 Db 897 CATTACCAATCGGGATCTTCACTGGA---AAGTGTGG-----ACAAATCTAGAT 944
 Qy 298 HisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIleProTyrTrpLeuLys 317
 Db 945 CATGCAGTGGTGGCTGTGGTTATGTTTCAGAGAACCGGTATTCATCTATTCGATTGTAAGG 1004
 Qy 318 AsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLysMetGluLeuGlyLysAsn 337
 Db 1005 AACTCGTGGGTACACGTTGGGGAGAGGATGTTTACATTAGGATGGAG-----AGAAC 1058
 Qy 338 Met-----CysGlyValAlaThrCysAlaSerTyrProile 349
 Db 1059 TTGCAAGTCCAAAGTCGGCAAGTGTGGAATTCGGTTCGATGAAGCTCGTACCCCGTT 1115

RESULT 13

US-08-821-994-82

; Sequence 82, Application US/08821994A

; Patent No. 6228643

GENERAL INFORMATION:

; APPLICANT: Greenland, Andrew J
 ; APPLICANT: Thomas, Didier RP
 ; APPLICANT: Jepson, Ian
 ; TITLE OF INVENTION: Promoters
 ; FILE REFERENCE: PPD 50108
 ; CURRENT APPLICATION NUMBER: US/08/821,994A
 ; CURRENT FILING DATE: 1997-03-22
 ; EARLIER APPLICATION NUMBER: PCT/GB97/00729
 ; EARLIER FILING DATE: 1997-03-18
 ; EARLIER APPLICATION NUMBER: GB 9606062.9
 ; EARLIER FILING DATE: 1996-03-22
 ; NUMBER OF SEQ ID NOS: 89

[illegible]

; SOFTWARE: PatentIn Ver. 2.1					
Db	384	TTATACCTCGGGCAAGAACCGGACCTGTCCGCCGCATCATTAANGGCCAAGAAGCGTAA	443		
Qy	130	mHisArgPheValaspGlyVal-----LeuProValThrArgAspTrpargGl	146		
Db	444	CATGAATAACTCAGCCGCGAGTAAACGACGTCGGAGGTTCCGGAGACGGTGACATGGAGACA	503		
Qy	146	uGlnGlyIleValSerProValLysAspGlnGlySerCysGlySer---	165		
Db	504	GAAAGGAGCGGTTAATGCCATTAAAGACCRAAGAACTTCGGGAAGTTGTTGGGGGTTTTTC	563		
Qy	165	rThrThrGlyAlaLeuGluAlaLaTyThrGlnLeuThrGlySer-----ThrLeuSe	183		
Db	564	AACAGCTCAGCAGTAGAAGGTATAAACCAAGATCGTAACAGAGAACTCATATCTCTGTCT	623		
Qy	183	rGlnGlnGlnLeuValLysAspCysAlaSerAlaPheAsnAsnPheGlyCys---GlyGlyLe	202		
Db	624	CGAACCAAGAACTTGTCTGACTCGGACAAATCATACACCAA---GGCTGTAAACGGCGTCT	680		
Qy	202	uProSerGlnAlaPheGluTyValLysTyAsnGlyGlyIleAspThrGluGlnThry	222		
Db	681	AATGGATTATGCTTTCAATCATCATGAAAAACGGCGGATTAACACCGAGAAAGACTA	740		
Qy	222	pProTyrrLeuGlyValMetGlyIleCysAsnPheLysGln--GluAsnValGlyValLys	241		
Db	741	TCTTTACCACCGAAACCAATGCAAAATGCAACTCTNNNTACTTTAAAATTTCAAGAGTTGTG	800		
Qy	242	VallileAspSerIle---AsnIleThrLeuGlyAlaGluAspGluLeuLysHisAlaVal	260		
Db	801	ACTATCGATGATACGAGATGTTCTTCTAGTAAGATGAACCGCGTTGAAGAGACGAGTT	860		
Qy	261	GlyLeuValArgProValSerValalaphGluVal---ValLysGlyPheAsnLeuTyx	279		
Db	861	TCG---TACCAGCCTGTGAGTGTGCTATTGATGCTGGTGAAGAGGTTTCCAACATTAC	917		
Qy	280	LysLysGlyValTyrrSerSerAspThrCysGlyArgAspProMetAspValAsnHisAla	299		
Db	918	CAATCTGGAACTTCACCTGGA---AAGTGTGCT-----ACGACTATGGATCACGCT	965		
Qy	300	ValleualaValGlyTyrrGlyValGluAspGlyIleProTyrrTpLeulleLysAsnSer	319		
Db	966	GTGCTGGCGGTGGTTATGTTTACAGAGAACCGGTGTGACTATTGGATTGTACGTAACCTCT	1025		
Qy	320	TrpGlyThrAsnTrpGlyAspAsnGlyTyrrPhelyMetGluLeuGly-----	335		
Db	1026	TGGGTACAGTTGGGGAGAGGATGTTACATTAGGATGGAGAGAAAGCGGCGNNGTCC	1085		
Qy	336	---LysAsnMetCysGlyValalaThrCysAlaSerTyrrProfile	349		
Db	1086	AAATCCGGTAAAGTGGGGATTCGGATTGAAGCCTCGTATCCGTT	1130		
RESULT 14					
US-08-821-994-86					
; Sequence 86, Application US/08821994A					
; Patent No. 6228643					
; GENERAL INFORMATION:					
; APPLICANT: Greenland, Andrew J					
; APPLICANT: Thomas, Didier RP					
; APPLICANT: Jepson, Ian					
; TITLE OF INVENTION: Promoters					
; FILE REFERENCE: PPD 50108					
; CURRENT APPLICATION NUMBER: US/08/821,994A					
; EARLIER FILING DATE: 1997-03-22					
; EARLIER APPLICATION NUMBER: PCT/GB97/00729					
; EARLIER FILING DATE: 1997-03-18					
; EARLIER APPLICATION NUMBER: GB 9606062.9					
; EARLIER FILING DATE: 1996-03-22					
; NUMBER OF SEQ ID NOS: 89					
; SOFTWARE: PatentIn Ver. 2.1					
; SEQ ID NO 86					
; TYPE: DNA					
; ORGANISM: Brassica napus					
US-08-821-994-86					

Db	384	TTATACCTCGGGCAAGAACCGGACCTGTCCGCCGCATCATTAANGGCCAAGAAGCGTAA	443
Qy	130	mHisArgPheValaspGlyVal-----LeuProValThrArgAspTrpargGl	146
Db	444	CATGAATAACTCAGCCGCGAGTAAACGACGTCGGAGGTTCCGGAGACGGTGACATGGAGACA	503
Qy	146	uGlnGlyIleValSerProValLysAspGlnGlySerCysGlySer---	165
Db	504	GAAAGGAGCGGTTAATGCCATTAAAGACCRAAGAACTTCGGGAAGTTGTTGGGGGTTTTTC	563
Qy	165	rThrThrGlyAlaLeuGluAlaLaTyThrGlnLeuThrGlySer-----ThrLeuSe	183
Db	564	AACAGCTCAGCAGTAGAAGGTATAAACCAAGATCGTAACAGAGAACTCATATCTCTGTCT	623
Qy	183	rGlnGlnGlnLeuValLysAspCysAlaSerAlaPheAsnAsnPheGlyCys---GlyGlyLe	202
Db	624	CGAACCAAGAACTTGTCTGACTCGGACAAATCATACACCAA---GGCTGTAAACGGCGTCT	680
Qy	202	uProSerGlnAlaPheGluTyValLysTyAsnGlyGlyIleAspThrGluGlnThry	222
Db	681	AATGGATTATGCTTTCAATCATCATGAAAAACGGCGGATTAACACCGAGAAAGACTA	740
Qy	222	pProTyrrLeuGlyValMetGlyIleCysAsnPheLysGln--GluAsnValGlyValLys	241
Db	741	TCTTTACCACCGAAACCAATGCAAAATGCAACTCTNNNTACTTTAAAATTTCAAGAGTTGTG	800
Qy	242	VallileAspSerIle---AsnIleThrLeuGlyAlaGluAspGluLeuLysHisAlaVal	260
Db	801	ACTATCGATGATACGAGATGTTCTTCTAGTAAGATGAACCGCGTTGAAGAGACGAGTT	860
Qy	261	GlyLeuValArgProValSerValalaphGluVal---ValLysGlyPheAsnLeuTyx	279
Db	861	TCG---TACCAGCCTGTGAGTGTGCTATTGATGCTGGTGAAGAGGTTTCCAACATTAC	917
Qy	280	LysLysGlyValTyrrSerSerAspThrCysGlyArgAspProMetAspValAsnHisAla	299
Db	918	CAATCTGGAACTTCACCTGGA---AAGTGTGCT-----ACGACTATGGATCACGCT	965
Qy	300	ValleualaValGlyTyrrGlyValGluAspGlyIleProTyrrTpLeulleLysAsnSer	319
Db	966	GTGCTGGCGGTGGTTATGTTTACAGAGAACCGGTGTGACTATTGGATTGTACGTAACCTCT	1025
Qy	320	TrpGlyThrAsnTrpGlyAspAsnGlyTyrrPhelyMetGluLeuGly-----	335
Db	1026	TGGGTACAGTTGGGGAGAGGATGTTACATTAGGATGGAGAGAAAGCGGCGNNGTCC	1085
Qy	336	---LysAsnMetCysGlyValalaThrCysAlaSerTyrrProfile	349
Db	1086	AAATCCGGTAAAGTGGGGATTCGGATTGAAGCCTCGTATCCGTT	1130
RESULT 14			
US-08-821-994-86			
; Sequence 86, Application US/08821994A			
; Patent No. 6228643			
; GENERAL INFORMATION:			
; APPLICANT: Greenland, Andrew J			
; APPLICANT: Thomas, Didier RP			
; APPLICANT: Jepson, Ian			
; TITLE OF INVENTION: Promoters			
; FILE REFERENCE: PPD 50108			
; CURRENT APPLICATION NUMBER: US/08/821,994A			
; EARLIER FILING DATE: 1997-03-22			

Db 894 GTGGCGTAAATCCGGTAAGTGTGGGATTGGGATAGAGCCCTCGTATCCGGT 947

RESULT 15
US-08-821-994-60
; Sequence 60, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier R
; APPLICANT: Jenson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: EPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 60
; LENGTH: 1553
; TYPE: DNA
; ORGANISM: Brassica napus
US-08-821-994-60
Alignment Scores:
Pred. No.: 1.69e-57 Length: 1553
Score: 557.00 Matches: 136
Percent Similarity: 54.87% Conservative: 61
Best Local Similarity: 37.88% Mismatches: 116
Query Match: 29.87% Indels: 46
Gaps: 13
DB:
US-10-087-714-2 (1-352) x US-08-821-994-60 (1-1553)
QY 5 LeuLeuPhePheLeuLeuPheLeuValSerAlaLeuSerValAlaLeuAlaGlyPheGlu 24
Db 74 ATCATTTCTTCTCCCTTCTTCTAGTTTCTTCTGAGTTCCCATCATCATCTCCGAT 133
QY 25 GluAspAsnProLeuArgSerValThrGlnArgProAspSerLeuGluProAlaLeu 44
Db 134 GACATCTCCGAGCTG-----TTGACCGTTGGTCCAGACACGGC 175
QY 45 GlyValLeuGlySerCysArgHisAlaPheHisAlaPheAlaArgPheAlaArgTyrGly 64
Db 149 -----TTGACCGTTGGTCCAGACACGGC 175
QY 65 LysSerTyrGlySerGluGluGluLeuLysLysArgPheGlyPheValGluAsnLeu 84
Db 176 AAAACGTACGCTTCGGAAGAAGAGAGACACACAGGATTGAAATCTTTAGAGACAAATCAC 235
QY 85 AlaPheLeu-----ArgSerThrAsnArgLysAspLeuSerTyrThrLeuGlyIleAsnGln 103
Db 236 GACTTCGTACACACACACACACCGCATCGTAACTTACTTACTTCTCTCTCTCAATGCC 295
QY 104 PheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAlaAlaGlnAsnCys 123
Db 296 TTCGGGATCTGACTCACCACAGTTCAGAGCCCTCTGCTCTTGACTCTCTGCTTCTTCA 355
QY 124 Ser-----AlaThrAlaHisGlyAsnHis---ArgPheValAspGlyValLeuProVal 140
Db 356 GCTCGTTGCTGGTGGTAAAGGAAAGTGTGAGAACGTTGGGGGCAAGATTCAGAT 415
QY 141 ThrArgAspTrpArgGluGluGlyIleValSerProValLysAspGlnGlySerCysGly 160
Db 416 TCTGTTGATTGGAGAGAAAGAGAGCTGTACTTAATGTCAAGATCAAGAGAGTGGGA 475
QY 161 Ser---TTPThrPheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGly 179
Db 476 GCGTGTGCTCTTCTCGGCGACTGGAGCAATGAAGGAAATCAACAGATGTGTACAGGA 535
QY 180 Ser-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPhe 197

Alignment Scores:
Pred. No.: 9.83e-58 Length: 1102
Score: 557.00 Matches: 127
Percent Similarity: 59.43% Conservative: 62
Best Local Similarity: 39.94% Mismatches: 99
Query Match: 29.87% Indels: 30
Gaps: 14
DB:
US-10-087-714-2 (1-352) x US-08-821-994-86 (1-1102)
QY 56 PheAlaArgPheAlaArgTyrGlyLysSerTyrGlySerGluGluLeu----- 73
Db 12 TACTTAAGATGCTCTTGGAGCAGCGGAAAGTAAACAGCAGCAACGCGTATTATCAAC 71
QY 74 -----LysLysArgPheGlyLeuPheValGluAsnLeuAlaPheLeu-----ArgSer 89
Db 72 CAACAGACGAAGATTCATAATTTCAAGAGCAACCTAAGATTTCATGATCTACCAAC 131
QY 90 ThrAsnArgLysAspLeuSerTyrThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrp 109
Db 132 GAGAACCAACAGACGCTACTTACAAGCTTGTCTAACCATATTGCTTAATCTCACTAAC 191
QY 110 GluGluPheArgThrAsnArgLeuGlyAla-----AlaGlnAsnCysSerAla 125
Db 192 GATCACTACCGGAGTTTATACCTCGGGGCAAGAACCGACCTGTCCGCGCATCACTAAG 251
QY 126 ThrAlaHisGlyAsnHisArgPheVal-----AspGlyValLeuProValThr 141
Db 252 GCCAAGAGCTTAACATGAATACTCAGCGGCACTAAACGAGTGGAGGTTCCGCTGACG 311
QY 142 ArgAspTrpArgGluGlnGlyLeuValSerProValLysAspGlnGlySerCysGlySer 161
Db 312 GTTGAAGTGGAGACAGAAAGCGGTTAATGCAATTAAGACCAAGGAACCTTGCAGGAGT 371
QY 162 ---TTPThrPheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGlySer 180
Db 372 TGTGGCGGTTTTCACAGCTGAGTAGAGGTATTAACAGATCGTACAGAGAA 431
QY 181 -----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPheGly 198
Db 432 CTGCTATCTTTGCGCAACAGAACTTGTGAGTGGCAAAATCGTACAAACAA---GGC 488
QY 199 Cys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyIleLeu 217
Db 489 TGTAAAGCGGCTTAATGAGTATGCTTTTCAATTCATTAATGAAACCGCGGATTAAC 548
QY 218 ThrGluGlnThrTyrProTyrLeuGlyValMetGlyIleCysAsnPheLysGlnGluAsn 237
Db 549 ACCGAGAAAGACTATCTTACACAGGACCAATGGCAATGCAACTCTTTTACTTAAGAT 608
QY 238 ValGlyValLysValLeuAspSerIle---AsnIleThrLeuGlyAlaGluAspGluLeu 256
Db 609 TCAAGAGTTGTAATCTGATGATAGTACGAAGATGTTCTTAGTAAAGATGAAACCGGTTG 668
QY 257 LysHisAlaValGlyLeuValArgProValSerValAlaPheGluVal---ValLysGly 275
Db 669 AAGAGACAGTTTCA---TACCAGCTGTGAGTGTGCTTATTCATGCTGCTGGAAGACT 725
QY 276 PheAsnLeuTyrLysLysGlyValTyrSerSerAspThrCysGlyArgAspProMetAsp 295
Db 726 TTCACACATTACCAATCTGAAATCTTCACTGGA---AAGTGTGCT-----ACGAAT 773
QY 296 ValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIleProTyrTrpLeu 315
Db 774 ATGATCACGCTGTGGTGGCGGTGTGTTATGGTTCAGAGAACCGGCTTGACTATTGGATT 833
QY 316 IleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLysMetGluLeuGly 335
Db 834 GTACGTAACCTCTGCGGTACACGTTGGGAGAGAGATGTTTACATTAGATGGAGAGAAC 893
QY 336 -----LysAsnMetCysGlyValAlaThrCysAlaSerTyrProLeu 349

```
Db      536  GATCTCATGACCTCTCTGACGAGAACTAATTGATTGTGACAAGTCCCTACACGAT--- 592
Qy      198  GlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGlyIle 216
Db      593  GGATGCAATGGTGTCTCATGCACTAGCTTTTCAATTGTCAATTAATAAACCATGGAATC 652
Qy      217  AspThrGlnGlnThrTyrProTyrLeuGlyValMetGlyIleCysAsnPhelLysGlnGlu 236
Db      653  GACACAGAGAAAGATTATCTTATCAAGAACGTGATGGCACCTGTAAAAAAGATAAGTTG 712
Qy      237  AsnValGlyValLysValIleAspSer---IleAsnIleThrLeuGlyAlaGluAspGlu 255
Db      713  AAAAGAAAGTTGTGACAATTGATGCTGCGGTAAATCAACACGACGAGAAAGCG 772
Qy      256  LeuLysHisAlaValGlyLeuValArgProValSerValAlaPhe---GluValLys 274
Db      773  TTACTGGAAGCTGTAGCG---GCTCAGCCAGTTAGTGTGGCATCTGTGGCAGCGAGAGA 829
Qy      275  GlyPheAsnLeuTyrLysGlyValTyrSerSerAspThrCysGlyArgAspProMet 294
Db      830  GCGTTTCAGCTATACTCTAAGGGAATATTCTCT-----GCCCATGTTCAACA 877
Qy      295  AspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIleProTyrTyr 314
Db      878  TCATTGGACCAACGCACTGCTCATCGTAGGATACGGTTTCACAGAACGGTGTGATTACTGG 937
Qy      315  LeuIleLysAsnSerTyrGlyThrAsnTyrGlyAspAsnGlyTyrPheLysMetGluLeu 334
Db      938  ATCGTGAAGAACTCTTGGGGAAGAGTTGGGATATGGATGGGTTTATGCACATGCCCGT 997
Qy      335  -----GlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIle 349
Db      998  AACACCGGCAACTCGGAAGGAGTATGTGGAATCAATATGCTCGCTTCGTATCCCATC 1054
```

Search completed: March 30, 2004, 00:11:56
Job time : 86 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 29, 2004, 23:30:35 ; Search time 330 Seconds
(without alignments)

3971.210 Million cell updates/sec

Title: US-10-087-714-2

Perfect score: 1865

Sequence: 1 MAAKLFFLLFLVSALV.....ELGKMGVATCASYPIVAV 352

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Xgapop 10.0 , Xgapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 2458946 seqs, 1861504846 residues

Total number of hits satisfying chosen parameters: 4917892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10087714/runat_26032004_113454_1057/app.query.fasta_1.519
-DB=Published Applications NA -OEMT=fastap -SUFFIX=rnbp -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blotum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTWT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10087714@cgn_1_164@runat_26032004_113454_1057
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Match	Length	ID	Description

1	1819	97.5	1071	14	US-10-087-714-1
2	1321.5	70.9	1413	12	US-10-425-114-14419
3	1320.5	70.8	1398	12	US-10-425-114-156
4	1319.5	70.8	1383	12	US-10-425-114-30654
5	1319.5	70.8	1392	12	US-10-425-114-5115
6	1310.5	70.3	1367	12	US-10-425-114-5145
7	1310.5	69.7	1407	14	US-10-219-220-107
8	1280	68.6	1344	12	US-10-425-114-10837
9	1280	68.6	1554	12	US-10-424-599-53151
10	1263	67.7	1282	9	US-09-770-445-20
11	1252	67.1	1183	12	US-10-425-114-2263
12	1249	67.0	1390	12	US-10-425-114-18625
13	1234	66.2	1333	12	US-10-425-114-18625
14	1148	61.6	1652	14	US-10-424-599-122829
15	852.5	45.7	830	12	US-10-425-114-3201
16	845.5	45.3	888	14	US-10-219-220-238
17	838	44.9	904	12	US-10-425-114-6228
18	828.5	44.4	813	12	US-10-425-114-35711
19	780.5	41.8	781	12	US-10-425-114-5160
20	759	40.7	1433	14	US-10-210-120-84
21	759	40.7	1490	9	US-09-925-302-35
22	759	40.7	1503	14	US-10-101-510-412
23	759	40.7	1503	14	US-10-252-157-19
24	691	37.1	1106	14	US-10-101-510-153
25	656.5	35.2	677	14	US-10-219-220-111
26	615	33.0	1392	9	US-09-938-842A-1372
27	615	33.0	1392	11	US-09-938-842A-1372
28	609	32.7	1776	14	US-10-219-220-236
29	598.5	32.1	1410	12	US-10-425-114-6299
30	594.5	31.9	2054	14	US-10-219-220-234
31	594	31.8	1600	14	US-10-219-220-98
32	592.5	31.8	1656	12	US-10-425-114-32817
33	590.5	31.7	1730	12	US-10-425-114-30186
34	590.5	31.7	1746	12	US-10-425-114-22346
35	590.5	31.7	1746	12	US-10-425-114-30222
36	590	31.6	1463	14	US-10-219-220-92
37	589	31.6	1392	15	US-10-618-644-6
38	589	31.6	1712	12	US-10-425-114-10971
39	589	31.6	1857	12	US-10-425-114-29538
40	584	31.3	1407	14	US-10-259-165-723
41	584	31.3	1568	12	US-10-425-114-5245
42	584	31.3	1576	12	US-10-425-114-1080
43	584	31.3	1631	12	US-10-425-114-26666
44	583.5	31.3	1570	12	US-10-425-114-33813
45	583.5	31.3	1575	12	US-10-425-114-33200

ALIGNMENTS

RESULT 1

US-10-087-714-1
: Sequence 1, Application US/10087714
: Publication No. US20030070188A1
: GENERAL INFORMATION:
: APPLICANT: Havkin-Frenkel, Daphna
: APPLICANT: Podstolski, Andrzej
: APPLICANT: Dixon, Richard A.
: TITLE OF INVENTION: Vanillin Biosynthetic Pathway Enzyme From Vanilla Planifolia
: FILE REFERENCE: DMC10099
: CURRENT APPLICATION NUMBER: US/10/087,714
: PRIOR FILING DATE: 2002-02-28
: PRIOR APPLICATION NUMBER: 09/462,576
: PRIOR FILING DATE: 2000-05-22
: PRIOR APPLICATION NUMBER: PCT/US98/14895
: PRIOR FILING DATE: 1998-07-15
: PRIOR APPLICATION NUMBER: 60/052,604
: PRIOR FILING DATE: 1997-07-15
: PRIOR APPLICATION NUMBER: 60/272,415
: PRIOR FILING DATE: 2001-02-28
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: Patent in version 3.1
: SEQ ID NO 1
: LENGTH: 1071

Sequence 1, Appli

Sequence 14419, A
Sequence 156, App
Sequence 30654, A
Sequence 5116, Ap
Sequence 5145, Ap
Sequence 107, App
Sequence 10837, A
Sequence 53151, A
Sequence 20, Appl
Sequence 2263, Ap
Sequence 18625, A
Sequence 122629, A
Sequence 241, App
Sequence 3201, App
Sequence 238, App
Sequence 6228, App
Sequence 35711, A
Sequence 5160, App
Sequence 84, Appl
Sequence 35, Appl
Sequence 412, App
Sequence 19, Appl
Sequence 153, App
Sequence 111, App
Sequence 1372, Ap
Sequence 1372, Ap
Sequence 236, App
Sequence 6299, App
Sequence 234, App
Sequence 98, Appl
Sequence 32817, A
Sequence 30186, A
Sequence 22346, A
Sequence 30222, A
Sequence 92, Appl
Sequence 6, Appl
Sequence 10971, A
Sequence 29538, A
Sequence 723, App
Sequence 5245, Ap
Sequence 1080, Ap
Sequence 26666, A
Sequence 33813, A
Sequence 33200, A

297 AsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIleProTyrTrpLeu 316
301 AACACGCGATTTCTTCCGCTCGGTTATGGAGTCGAGGACGGGATTCCTTATGGCTCATC 960
317 LysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLysMetGluLeuGlyLys 336
361 AAGAACTCATGGGGTGACAAATGGGGTGACAAATGGCTACTTTAAGATGGAATCGGCAAG 1020
337 AsnMetCysGlyValAlaThrCysAlaSerTyrProIleValAlaVal 352
1021 AACATGTGTGGTGTTCACACTTGGCATCTTATCCACTTGTGGCTGTG 1068

RESULT 2
US-10-425-114-14419
; Sequence 14419, Application US/10425114
; Publication No. US20040034889A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 14419
; LENGTH: 1413
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB189-021-F2_FLI
US-10-425-114-14419

Alignment Scores:
Pred. No.: 2.58e-163 Length: 1413
Score: 1321.50 Matches: 258
Percent Similarity: 79.17% Conservative: 27
Best Local Similarity: 71.67% Mismatches: 66
Query Match: 70.86% Indels: 9
DB: 12 Gaps: 6

US-10-087-714-2 (1-352) x US-10-425-114-14419 (1-1413)

QY 1 MetAlaAlaLysLeuLeuPhePheLeuPheLeuValSerAlaLeuSerValAlaLeu 20
DB 76 ATGCCCCACGCCGCTGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 135
QY 21 Ala-----GlyPheGluGluAspAsnProIleArgSerValThrGlnArgProAspSer 38
DB 136 GCCAACTCCGGCTTCGGGATCCAAACCAGATCCGCCCTCCCGCTACCGACCGCGGCGCTCC 195
QY 39 ---IleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
DB 196 GCGCTCGAGTCCACGGTCTTCGCGCGCTTCGCGCGCACCCGCGACGCTCCGCTTCGCA 255
QY 58 ArgPheAlaArgArgTyrGlyLysSerTyrGlySerGluGluIleLysLysArgPhe 77
DB 256 CGCTTGTGTGAGTACGGCAAGAGCTACGAGACCGCGCGGAGGTCCACAAAGCGGTT 315
QY 78 GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyr 97
DB 316 AGGATCTTCTCCGAGAGTCTCCAGCTCGCTCGCTCCACCAATCGGAAAGGCTCTCCTAC 375
QY 98 ThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeu 117
DB 376 CGCTTCGCGCATCAACCGTTCGCGACATGATCGGAGAGGTTCGTCGCGACCGGCTC 435
QY 118 GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPhe-----ValAsp 135

```

; TYPE: DNA
; ORGANISM: Vanilla planifolia
; US-10-087-714-1

Alignment Scores:
Pred. No.: 5,27e-329 Length: 1071
Score: 1819.00 Matches: 349
Percent Similarity: 98.60% Conservative: 2
Best Local Similarity: 98.03% Mismatches: 1
Query Match: 97.53% Indels: 4
DB: 14 Gaps: 3

US-10-087-714-2 (1-352) x US-10-087-714-1 (1-1071)

QY 1 MetAlaAlaLysLeuLeuPhePheLeuLeuPheLeuValSerAlaLeuSerValAlaLeu 20
DB 1 ATGGCAGGTAAGCTCCCTCTCTCTACTCTTCTGGTCTCGGCCCTCTCGGTGCGGCTC 60
QY 21 AlaGlyPheGluGluAspAsnProIleArgSerValThrGlnArgProAspSerIleGlu 40
DB 61 GCGCGGTTTCGAAGAGAGCAATCCCAATCCGCTCGGTATACAAAGGCGCTGACTCGGATTGAG 120
QY 41 ProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAla 60
DB 121 CCTGCATCCTCGGGTCTCTGGCAGTTGCGGCACGCGCTTCACATTCGCACGGTTGCC 180
QY 61 ArgArgTyrGlyLysSerTyrGlySerGluGluGluIleLysLysArgPheClyIlePhe 80
DB 181 CGCAGGATACGGAGAGAGCTACGGATCGGAGGAGGAGATCAAGAAAGAGGTTCGGGATCTTC 240
QY 81 ValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyrThrLeuGly 100
DB 241 GTGGAGGAATCTAGCGTTTATCCGCTCCACTAATCGGAAGGATCTGTCGTATACCTAGGA 300
QY 101 IleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAlaAla 120
DB 301 ATCAACCAATTCGCGACCTCAGCTCGGAGGAATTCGGACCAATCGCCTTCGGTCGGCG 360
QY 121 GlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProVal 140
DB 361 CAGAACTGCTCGGCGACTCGGCATCGAAACACCGGTTTGTTCGATGGCGTGTCTCTGTA 420
QY 141 ThrArgAspTrpArgGluGluGlyIleValSerProValLysAspGlnGlySerCysGly 160
DB 421 ACGAGGGATTTGAGGAGGACCAAGGATAGTAGGCCCTGTAAAGACCAAGAAAGCTGTGGA 480
QY 161 Ser---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGly 179
DB 481 TCTTGCTGGACCTTTCACTACTCTGGACACTAGAGGCTGCATATACACAGCTAACTGGA 540
QY 180 -----SerThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsn 197
DB 541 AAGAGCAGCATCATTTATCTGAACAGCAACTGTGGACTGTGCCCTCAGCATTCATTAACCTT 600
QY 198 GlyCys--GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGlyIle 216
DB 601 GGAATGCAATGGAGGTTTGCCTTCCCAAGCCCTTTGAATACGTTAAGTACAATGGAGGCATC 660
QY 217 AspThrGluGlnThrTyrProTyrLeuGlyValMetGlyIleCysAsnPheLysGlnGlu 236
DB 661 GACACAGAACAGACTTATCCATACCTCGGTGTCAATGGTATCTGCAACTTCAGACAGGAG 720
QY 237 AsnValGlyValLysValIleAspSerIleAsnIleThrLeuGlyAlaGluAspGluLeu 256
DB 721 AATGTTGGTGTCAAGGTCATTGATTCGATAAATCATCACCCCTGGGTGCTGAGGATGAGTTG 780
QY 257 LysHisAlaValGlyLeuValArgProValSerValAlaPheGluAlaValLysGlyPhe 276
DB 781 AAGCATGCAATGGGCTTTGGTGGCTCCAGTTAGCGTTGCATTTGAGGTTGTGAAGGTTTC 840
QY 277 AsnLeuTyrLysLysGlyValTyrSerSerAspThrCysGlyArgAspProMetAspVal 296
DB 841 AATCTCTCAAGAGGCTATATACAGAGTGAACCTCTGTGGAAGAGATCCAAATGGATGTCG 900

```



```

436 GCGCAGCCAGAACTGCTCCGCCACGCTTACCGCAACACCGGATGACGCGCGCGCC 495
136 GlyValLeuProValThrArgAspTrpArgGluGlnGlyLeuValSerProValLysAsp 155
496 GTTCGGCTGCCGAGAGCAAGAGCTGGAGGAGGATGGGATTTGGACCCAGTGAAGAAC 555
156 GlnGlySerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaTyr 174
556 CAGGCCCACTGTGGATCATGTGACCTTACGACACTACTGGCGCCTTGGAGGTGCATAT 615
175 ThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSer 192
616 ACCCAGGCAACTGGCAAGCCCATCTCTCTCAGCAACAGCTTGTGTGACTGTGGTTT 675
193 AlaPheAsnAsnGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLys 211
676 GCATTCAACAATTTCCGATGCAACGAGGCTTCCATCCCGGCTTTGTAATACATCAA 735
212 TyrAsnGlyGlyLeuAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyLeuCys 231
736 TACAATGGTGGCTTGCACCTGAGCAATCTTACCCCTTACCAAGGTGTCAATGGAATCTGC 795
232 AsnPhelysGlnGluAsnValGlyValLysValLeuAspSerLeuAsnLeuThrLeuGly 251
796 AAGTTTAAAGATGAGATGTTGGAGTCAAGGTTTGGACTCGGTTAATACATCACCTGGGT 855
252 AlaGluAspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGlu 271
856 GCTCAGGATGAATGAAGGATGCTGTTGGTCTGGTTCGCTGCTGCTGCTGCTGCTGCTG 915
272 ValValLysGlyPheAsnLeuTyrLysGlyValTyrSerSerAspThrCysGlyArg 291
916 GTGATCACTGTTTTCAGGCTGTACAGAGCGGAGTTTACACTAGCGACCACTGTGGAAT 975
292 AspProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyLe 311
976 ACACCGATGATGAACACCGCTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1035
312 ProTyrTrpLeuLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLys 331
1036 CCTTACTGGCTTCATCAAGAACTCATGGGGCGCTGACTGGGGTGATGATGTTACTTCAAG 1095
332 MetGluLeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProLeuValAla 351
1096 ATGGAATGGCAGACATGTCGGTGTGCTACGTGTGCTACCTTACCTTATGTCGCA 1155

```

RESULT 3

```

US-10-425-114-156
; Sequence 156, Application US/10425114
; Publication No. US20040034889A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 156
; LENGTH: 1398
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700048693_FLI
US-10-425-114-156

```

Alignment Scores:

```

Pred. No.: 3,44e-163 Length: 1398
Score: 1320.50 Matches: 258
Percent Similarity: 78.89% Conservative: 26
Best Local Similarity: 71.67% Mismatches: 67
Query Match: 70.80% Indels: 9
DB: 12 Gaps: 6

US-10-087-714-2 (1-352) x US-10-425-114-156 (1-1398)

QY 1 MetAlaAlaLysLeuLeuPhePheLeuLeuValSerAlaLeuSerValAlaLeu 20
DB 68 ATGGCCCAACGCGCGCTGCTGCTCTCGCGCTGCTGCGCCCTCGCGCCACCGCGCGCGG 127
QY 21 Ala-----GlyPheGluGluAspAsnProLeuArgSerValThrGluArgProAspSer 38
DB 128 GCCAACTCCGGCTTCGCGGACTCCAAACCGATCCCGCCCGCTCAGCGACCGCGCGCTCC 187
QY 39 ---LLeuProAlaLLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
DB 188 GGCCTCGAGTCCACGCTCTTCGCGCTGCGCCCGCACCCCGGACGCGCTCCGCTTCGCA 247
QY 58 ArgPheAlaArgTyrGlyLysSerTyrGlySerGluGluGluLeuLysLysArgPhe 77
DB 248 CGCTTCGCGCTCAGGTACGGCAAGAGTACGAGAGCGCGCGGAGGTCACACAAGCGGTTC 307
QY 78 GlyLLeuPheValGluAsnLeuAlaPheLeuArgSerThrAsnArgLysAspLeuSerTyr 97
DB 308 AGGATCTTCTCCGAGAGCTCCAGCTGCTCGCTCCCAACCGCAAGGCGCTCTCCCTAC 367
QY 98 ThrLeuGlyLysAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeu 117
DB 368 CGCTTCGCGCATCAACGCTTCGCGCATGAGTGGAGGAGTTCGCTGCGACCGCGCTC 427
QY 118 GlyAlaAlaGluAsnCysSerAlaThrAlaHisGlyAsnHisArgPhe-----ValAsp 135
DB 428 GCGCAGCCCAAGAACTGCTCGCCACGCTTACCGCAACACCGGATGCGCGCCCGCC 487
QY 136 GlyValLeuProValThrArgAspTrpArgGluGlnGlyLysValSerProValLysAsp 155
DB 488 GTTCGGCTGCCGAGACCAAGAGCTGGAGGAGGATGGATTTGAGCCCGAGTGAAGAAC 547
QY 156 GlnGlySerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaTyr 174
DB 548 CAGGGCCACTGTGGATCATGCTGGACCTTCAGCACTACTGTGTGCTGCTGAGGCTGCATAT 607
QY 175 ThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSer 192
DB 608 ACCAGGCAACTGGCAAGCCCATCTCTCTGAGCAACAGCTTGTGACTGTGCTTTT 667
QY 193 AlaPheAsnAsnGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLys 211
DB 668 GCATTCAACAATTTCCGATGCAACGAGGCTTCCATCCCGGCTTTGTAATACATCAA 727
QY 212 TyrAsnGlyGlyLeuAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyLeuCys 231
DB 728 TACAATGGTGGCTTGCACACTGAGAACTTACCTTACCAAGGTTGTAATGGAATTCG 787
QY 232 AsnPhelysGlnGluAsnValGlyValLysValLysAspSerLeuAsnLeuThrLeuGly 251
DB 788 AAGTTTAAAGATGAGATGTTGGAGTCAAGTTTTGGACTCGGTTAATACATCACCTGGGT 847
QY 252 AlaGluAspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGlu 271
DB 848 GCTGAGGATGAATGAAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 907
QY 272 ValValLysGlyPheAsnLeuTyrLysGlyValTyrSerSerAspThrCysGlyArg 291
DB 908 GTGATCACTGTTTTCAGGCTGTACAGAGCGGAGTTTACACTAGCGACCAATTTGGA 967
QY 292 AspProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyLe 311
DB 968 ACACCGATGATGAACCAACGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1027

```

```

312 ProTyrTrpLeuIleLysAsnSerTyrPglYThrAsnTrpClyAspAsnGlyTyrPhelys 331
1028 CCTACTGGCTCATCAAGAACTCATGGGGGGCTGACTGGGGCGATGAGGGTTACTTCAAG 1087
332 MetGluLeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIleValAla 351
1088 ATGGAAATGGGCAAGAACATGTGCGGTGTTGCTACGTGTGCATCTTACCTATTGTGCA 1147

RESULT 4
US-10-425-114-30654
; Sequence 30654, Application US/10425114
; Publication NO. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313) B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 30654
; LENGTH: 1383
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-2MFLB73058A12_FLI
US-10-425-114-30654

```

Alignment Scores:		
Pred. No.:	4.58e-1c3	Length:
Score:	1319.50	Matches:
Percent Similarity:	78.89%	Conservative:
Best Local Similarity:	71.39%	Mismatches:
Query Match:	70.75%	Indels:
DB:	12	Gaps:
		6
US-10-087-714-2 (1-352) x US-10-425-114-30654 (1-1383)		

[illegible]

475 GTTCCGCTCCGGAGACGAAAGACTGGAGGGAGATGGGATTTGTGAGCCCGAGTGAAAAAC 534
 156 GlnGlySerCysGlySer---TrrThrPheserThrGlyAlaLeuGluAlaAlaTyr 174
 535 CAGGGCCACTGTGGATCATGCTGGACACTTCCAGCACACTACTGGTGCACTTGGAGGCTGCATAT 594
 175 ThrGlnLeuThrGlySer-----ThrLeuSerGlnGlnGlnLeuValAspCysAlaSer 192
 595 ACCCAGGCAACTGGCAAGCCCATCTCTCTCTGAGCAACAGCTTATTGACTGTGGTTTT 654
 193 AlaPheAsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLys 211
 655 GCATTCAACAATTTCCGATGCAACCGAGGCGCTTCCATCCAGGCGCTTTGGAATACATCAAA 714
 212 TyrAsnGlyClyIleAspThrGlnGlnThrTyrProTyrLeuGlyValMetGlyIleCys 231
 715 TACAATGGTGGCTGTGACATCGAGGAGCTTACCTTTCCCAAGGTGTCAATGGAACTCTGC 774
 232 AsnPhelysGlnGluAsnValGlyValLysValIleAspSerIleAsnIleThrLeuGly 251
 775 AAGTTTAAGAAATGAGAATGTTGGAGTCAAGGTTTTGGACTCGGTTAAACATCACCCCTGGT 834
 252 AlagluAspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGlu 271
 835 GCTGAGGATGAACCTGAAGGATGCTGTGGTCTGGTTCGCCAGTTAGTTGGCTTCGAG 894
 272 ValValLysGlyPheAsnLeuTyrLysGlyValTyrSerSerAspThrCysGlyArg 291
 895 GTGATCACTGGTTTCAGGCTGTACAGAGCGGAGTTTACATAGCGACCATTTGTGGAAC 954
 292 AspProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIle 311
 955 ACACCGATGGATGTGAACACCGCTGTTCTCGCTGTGGCTACGGTGTGCGAGATGGTGT 1014
 312 ProTyrTrpLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPhelys 331
 1015 CCTTACTGGCTTCATCAAGAACTCATGGGGCGCTGACTGGGGTGATGAGGTTTACTTCAAG 1074
 332 MetGluLeuGlyLysAsnMetCysGlyValAlaAlaThrCysAlaSerTyrProIleValAla 351
 1075 ATGGAATGGCAAGAACATGTGGCGTGTGTACGTGTGATCCTACCTATTGTGCGCA 1134
 RESULT 5
 US-10-425-114-5116
 ; Sequence 5116, Application US/10425114
 ; Publication No. US20040034888A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Jingdong
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Screen, Steven E
 ; APPLICANT: Tabaska, Jack E
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53313)B
 ; CURRENT APPLICATION NUMBER: US/10/425,114
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 73128
 ; SEQ ID NO 5116
 ; LENGTH: 1392
 ; TYPE: DNA
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: 700443856_FLI
 US-10-425-114-5116
 Alignment Scores:
 Pred. No.: 4,63e-163 Length: 1392
 Score: 1319.50 Matches: 257
 Percent Similarity: 78.89% Conservative: 27
 Best Local Similarity: 71.39% Mismatches: 6
 Query Match: 70.75% Indels: 9

Alignment Scores:		
Pred. No.:	4,635-163	1392
Score:	1319.50	257
Percent Similarity:	78.8%	Matches:
Best Local Similarity:	71.3%	Conservative: 27
Query Match:	70.7%	Mismatches: 67
		Indels: 9

DB: 12 Gaps: 6

US-10-087-714-2 (1-352) x US-10-425-114-5116 (1-1392)

Qy 1 MetAlaLysLeuLeuPhePheLeuPheLeuValSerAlaLeuSerValAlaLeu 20

Db 55 ATGGCCCAACCGCGCTTGCTGCTGCGGTGCTGCGCCCTCGCGGCCACCGCGCGCGG 114

Qy 21 Ala-----GlyPheGluGluAspAsnProIleArgSerValThrGlnArgProAspSer 38

Db 115 GCCAACTCCGGCTTCGGGGACTCAACCCGATCGCCCGCGCTCAACCGACCGCGCGCGCTCC 174

Qy 39 --rIleGlnProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57

Db 175 GCGCTCAGATCCACGGCTTCGCGCGGCTCGCGCGACCCCGCGACGCGCTCGCTTCGCA 234

Qy 58 ArgPheAlaArgArgTyrGlyLysSerTyrGlySerGluGluGluIleLysLysArgPhe 77

Db 235 CGCTTCGCTGTGAGGTACGGCAAGAGCTACGAGAGCGCGCGGAGGTCCACAAAGCGGTTC 294

Qy 78 GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyr 97

Db 295 AGGATCTTCCGAGAGTCTCCAGCTCGTCGCGCTCCACCAATCGGAAGGCGCTTCCTCAC 354

Qy 98 ThrLeuGlyIleAsnGlnPheAlaAspLeuThrTyrGluGluPheArgThrAsnArgLeu 117

Db 355 CGCCTCGCATCAACCGCTTCGCGCATGAGCTGGGAGGAGTCCGTGGACCCGCGCTC 414

Qy 118 GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPhe-----ValAsp 135

Db 415 GGGCGACCCAGAACTCGTCGCCACACGCTTACCGGGCAACACCCGGATGCGCGCGCGGCC 474

Qy 136 GlyValLeuProValThrArgAspTyrArgGluGlnGlyIleValSerProValLysAsp 155

Db 475 GTTCGGCTCGCGAGAGCAAGACTGGAGGAGGATGGATTGTGAGCCCAAGTGAANAAC 534

Qy 156 GlnGlySerCysGlySer---TyrThrPheSerThrThrGlyAlaLeuGluAlaAlaTyr 174

Db 535 CAGGGCCACTGTGTGATCATGTGACCTTCAGCACTACTGTGTGCATTTGAGGCTGCATAT 594

Qy 175 ThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSer 192

Db 595 ACCCAGGCACTGGCAAGCCCATCTCTCTCTGAGCAACAGCTATTATGACTGTGGTTTT 654

Qy 193 AlaPheAsnAsnPheGlyCys--GlyGlyLeuProSerGlnAlaPheGluTyrValLys 211

Db 655 GCATTTCAACAATTCGGATGCAACGGAGGCGCTTCCATCCCGAGGCTTTGAATACATCAA 714

Qy 212 TyrAsnGlyGlyIleAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyIleCys 231

Db 715 TACAATCGTGCGCTTGACCTAGAGAAATCTTACCCTTTACCAAGGTGTCAATGGAAATCGC 774

Qy 232 AsnPheLysGlnGluAsnValGlyValLysValIleAspSerIleAsnIleThrLeuGly 251

Db 775 AAGTTTAAAGATGAGATGTTGGAGTTTGGACTCGGTAACTACATCACTCACCTCGGT 834

Qy 252 AlaGluAspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGlu 271

Db 835 GCTCAGGATGAACCTGAAGAGT 894

Qy 272 ValValLysGlyPheAsnLeuTyrLysLysGlyValTyrSerSerAspThrCysGlyArg 291

Db 895 GTGATCTGTTTCAGGCTGTACAGAGCGAGTTTACATCAGCACCATTTGTGGAACT 954

Qy 292 AspProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIle 311

Db 955 ACACCGATGGATGTGAACCGCTGTTCTGGCTGTGGTGTGGTGTGGTGTGGTGTGGTGTGTA 1014

Qy 312 ProTyrTyrPheLysAsnSerTyrGlyThrAsnTyrGlyAspAsnGlyTyrPheLys 331

Db 1015 CCCTACTGGCTATCAAGAACTCATGGGCGCGCTCACTGGGGGTGATGAGGGTTACTTCAG 1074

Qy 332 MetGluLeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIleValAla 351

```

Db      1075  ATCGAAATGGCAAGCAATGTGCGGTGTTGCTACGTGTGCATCTACCTCTATTGTGCGCA 1134
RESULT 6
US-10-425-114-5145
; Sequence 5145, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 5145
; LENGTH: 1367
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700444833_FLI
US-10-425-114-5145

Alignment Scores:
Pred. No.: 6,89e-162 Length: 1367
Score: 1310.50 Matches: 252
Percent Similarity: 81.18% Conservative: 24
Best Local Similarity: 74.12% Mismatches: 57
Query Match: 70.27% Indels: 7
DB: 12 Gaps: 5

US-10-087-714-2 (1-352) x US-10-425-114-5145 (1-1367)

Qy 19 AlaLeuAlaGlyPheGluGluAspAsnProIleArgSerValThrGlnArgProAspSer 38
Db 1 GCACATCTCGGCTTCGGGACTCCACCCGATCGCCCTGTCCACCGCGCGGCGCTCC 60
Qy 39 ---IleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
Db 61 CGCTCGAGTCCACGGCTCTTCGCCGCGTTCGCCGCGCACCCGCGAGCGGCTTCGCTTCGCA 120
Qy 58 ArgPheAlaArgArgTyrGlyValSerTyrGlySerGluGluGluIleLysLysArgPhe 77
Db 121 CGTTCCGCGTCAGTTACGCGAAGAGCTACGAGAGCGCGCGGAGGTCCATAGCGGTTTC 180
Qy 78 GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyr 97
Db 181 AGGATCTTCCGAGAGCGCTCCAGTGTGTCGCTCCACCAACGCAAGGCGCTCTCCTAC 240
Qy 98 ThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeu 117
Db 241 CGCCTCGGATCAACCGCTTCGCGACATGAGCTGGAGAGATTCCGTGGACCGCGCTC 300
Qy 118 GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPhe-----ValAsp 135
Db 301 GCGCGACGCCAGAACTGCTCCGCCACGTTTACCGCAACACCGAGTGC GCGCGCGCGC 360
Qy 136 GlyValLeuProValThrArgAspTrpArgGluGlnGlyIleValSerProValLysAsp 155
Db 361 GTTGCCTCCCGAGACCAAACTGGAGGAGGATGGGATTGTGAGCCAGTGAATAAC 420
Qy 156 GlnGlySerCysGlySer---TriPThrPheSerThrThrGlyAlaLeuGluAlaIaTyr 174
Db 421 CAGGGCCACTGTGGATCATGTGTGGACCTTCAGCACTACTGTGTGCATCTGAGGTGCATAT 480
Qy 175 ThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSer 192
Db 481 ACCCAGGCACTGGGAAGCCCATCTCTCTCTGAGCAACAGCTGTGTGACTGTGGTTTT 540

```

QY 193 AlaPheAsnAspGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrVallys 211
 Db 541 GCATTCAAAATTTCCGATGCAACGGAGGCTTCCATCCCGGCTTTGAATACATCAA 600
 QY 212 TyrAsnGlyGlyLeuAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyLeuCys 231
 Db 601 TACAATGGTGGCTTGACACTGAGGAATCTTACCCCTTACCAAGGTGTCAATGGAATCTGC 660
 QY 232 AsnPheLysGlnGluAsnValGlyVallysValIleAspSerIleAsnIleThrIleuGly 251
 Db 661 AAGTTTAAAGATGAGAATGTTGGAGTCAAGGTTTTGGAGTCCGTTAAACATCACCCCTGGGT 720
 QY 252 AlaGluAspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGlu 271
 Db 721 GCTGAGGATGAATCAAGATCTGTTGGTCTGGTTCGCCAGTAGGTTGCCCTCGAG 780
 QY 272 ValValLysGlyPheAsnLeuTyrLysLysGlyValTyrSerSerAspThrCysGlyArg 291
 Db 781 GTGATCACTGGTTTTCAGGCTGTACAGAGCGGAGTTTACACTAGCGACCATTTGTGGAAT 840
 QY 292 AspProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyLeu 311
 Db 841 ACACCGATGGATGGAACACACCTGTTCTGGCTGTTGGCTACGGTTCGAAGATGGTGTG 900
 QY 312 ProTyrTyrLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLys 331
 Db 901 CCTACTGGCTCATCAAGAACTCATGGGCGCTGACTGGGGCGATGAGGTTACTTCAAG 960
 QY 332 MetGluLeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIleValAla 351
 Db 961 ATGGAAATGGCAAGAACATGTCGGTGTGTGTACGTGTGATCTACCCCTATTGTGCA 1020

RESULT 7

US-10-219-220-107
 ; Sequence 107, Application US/10219220
 ; Publication No. US20030082724A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Flinn, Barry
 ; TITLE OF INVENTION: Compositions affecting programmed cell
 ; TITLE OF INVENTION: death and their use in the modification of plant development
 ; FILE REFERENCE: 11000.1022c1
 ; CURRENT APPLICATION NUMBER: US/10/219,220
 ; CURRENT FILING DATE: 2002-08-14
 ; PRIOR APPLICATION NUMBER: U.S. No. US20030082724A1 09/325,932
 ; PRIOR FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 290
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 107
 ; LENGTH: 1407
 ; TYPE: DNA
 ; ORGANISM: Eucalyptus grandis
 US-10-219-220-107

Alignment Scores:

Pred. No.: 1496-160 Length: 1407
 Score: 1300.50 Matches: 247
 Percent Similarity: 79.77% Conservatives: 33
 Best Local Similarity: 70.37% Mismatches: 60
 Query Match: 69.73% Indels: 11
 DB: 14 Gaps: 5

US-10-087-714-2 (1-352) x US-10-219-220-107 (1-1407)

QY 9 LeuLeuPheLeuValSerAlaLeuSerValAlaLeuAlaGlyPheGluGluAspAsnPro 28
 Db 91 CTCCTCTCGTCCGCTGCTGCTCTCTCCGCGCGGCTGAGGCTTCGAGAGTCCAAACCCC 150
 QY 29 IleArgSerValThrGlnArgProAsp-----SerIleGluProAlaIleLeu 44
 Db 151 ATCCGCTCTTC-----CCCGAGCGGCGCTCCGCGACCTCGAGTCTCCATCTGTC 201

RESULT 8

US-10-425-114-10837
 ; Sequence 10837, Application US/10425114
 ; Publication No. US20040034888A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Jingdong
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Screen, Steven E

QY 45 GlyValLeuGlySerCysArgHisAlaPheHisAlaArgPheAlaArgTyrGly 64
 Db 202 CAGATCGTGGCGCGCACCGCCACGCGCTTCTCTTCCGCGCTTCCGCAACAGGTATGG 261
 QY 65 LysSerTyrGlySerGluGluGluLysLysArgPheGlyIlePheValGluAsnLeu 84
 Db 262 AAGAGGTACGAGACCGCGGAGGAGATCAAGCTCGCGTTCGAGATCTTCAGGAGATCTC 321
 QY 85 AlaPheIleArgSerThrAsnArgLysAspLeuSerTyrThrLeuGlyIleAsnGlnPhe 104
 Db 322 AAGTTGATCGATCCACCAACAGAGGCGTTCCTTACCCCTCGTCTCAATAGTTT 381
 QY 105 AlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAlaAlaGlnAsnCysSer 124
 Db 382 GCTGATTGAGCTGGGAGGAGTTTCAAGAGCGCACAGCTGGAGCTCTCAAACTCTCT 441
 QY 125 AlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProValThrArgAspTyr 144
 Db 442 GCACACCAACAGGCGCACCAACAGCTCCAGGAGAGCTTCTCCCGAGATGAAGACTGG 501
 QY 145 ArgGluGlnGlyIleValSerProValLysAspGlnGlySerCysGlySer---TrpThr 163
 Db 502 AGAGAAAAGGCGATTGTAAGCCCAATTAAGATCAGGGCGACTGTGATCTTGTCTGACT 561
 QY 164 PheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGlySer-----Thr 181
 Db 562 TTCAGTACCACTGAGCTCTTGGAGTGTCTTATCACAAGCATTCGGGAAACAAATCTCT 621
 QY 182 LeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPheGlyCys---Gly 200
 Db 622 CTCTCTGAGCAGCAGCTCTGAGCTGTGGGCTTTCACAACTTTCGATGTAGTGGT 681
 QY 201 GlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGlyIleAspThrGluGln 220
 Db 682 GGACTGCCATCCCAAGCCTTGGATACGTCAAGTACACGGTGGCTTATATACCGAGGAA 741
 QY 221 ThrTyrProTyrLeuGlyValMetGlyIleCysAsnPheLysGlnGluAsnValGlyVal 240
 Db 742 GCATATCTTATACCGCAGTGGATGAGTAGTGCATAATTTCTCGCTGATATGTGTGTC 801
 QY 241 LysValIleAspSerIleAsnIleThrLeuGlyAlaGluAspGluLeuLysHisAlaVal 260
 Db 802 CAAGTGTCTGACTCTGTATACATCACCTTGGTGGTGGTGGATGAACATAAGCATGCGAT 861
 QY 261 GlyLeuValArgProValSerValAlaPheGluValValLysGlyPheAsnLeuTyrLys 280
 Db 862 GCCTTCTCGCGCAGTGGATGAGTGGCATTTCCAGTCTGTGAAGACTTCAGATTGTACAG 921
 QY 281 LysGlyValTyrSerSerAspThrCysGlyArgAspProMetAspValAsnHisAlaVal 300
 Db 922 TCGGTGTCTACACGAGCGATACATCGGTAGCATCTCCATGATGATGTAACCATGTCTT 981
 QY 301 LeuAlaValGlyTyrGlyValGluAspGlyIleProTyrTrpLeuIleLysAsnSerTyr 320
 Db 982 CTCGCTGTGTTGATGAGTTGAAGATGGTGTTCCTCTTGGCTCATCAGAAATTTCTGG 1041
 QY 321 GlyThrAsnTrpGlyAspAsnGlyTyrPheLysMetGluLeuGlyLysAsnMetCysGly 340
 Db 1042 GGAGCAGACTGGGTGACCGAGTACTTCAAGATGGAGATGGAAAGAACATGTGTGGA 1101
 QY 341 ValAlaThrCysAlaSerTyrProIleValAla 351
 Db 1102 GTCGCTACTTGTGCATCATACCTCTGTGTGGCC 1134


```

Db      372 ATATTTCGCAAAACCTCAGTTTCCTCCCTCCCAACAGAACCGCTTGCCCTACT 431
QY      99  LeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGly 118
Db      432 CTCTCTGTTAATCATTTTCTGATTGGACTTGGGAGGAGTTCAAAAGACACAGACTAGGC 491
QY     119 AlaaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeu 138
Db     492 GTTGCCCAAAATGCTCTCCACTCTTACGGCAACCAACCAAGCTCACTGATGCTGTTCT 551
QY     139 ProValThrArgAspTrpArgGluGlnGlyIleValSerProValIleAspGlnGlySer 158
Db     552 CTCCACCAAGAGACTGGAGGAAGAGAAATAGTGGTGGTAAAGATCAAGGCAGC 611
QY     159 CysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaTyThrGlnLeu 177
Db     612 TCGGATCATGTGGACATTCAGCACTGAGCGGCTTTAGAACGACGCTATGCACAGCA 671
QY     178 ThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsn 195
Db     672 TTTGGGAAGAGATCTCTCTTTCTGACGACAGCTAGTGGATGTGCTGGCCCTTCAAC 731
QY     196 AsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyVal-LysTyAsnGln 214
Db     732 AACTTTGGCTGTAATGCTGGGTGTCATCACAAGCCTTTGAGTACATTTAAATAACAATGG 791
QY     214 yGlyIleAspThrGluGlnThrTyProTyThrLeuGlyValMetGlyIleCysAsnPheTy 234
Db     792 TGACTAGACAGACAGAGAACATATCCCTACACAGGAAGAGATGGTGTCTGCAAAATCTC 851
QY     234 sGlnGluAsnValGlyValIleAspSerIleAsnIleThrLeuGlyValAlaGluAs 254
Db     852 AGCTGAAATGTTGCTGTTCAAGTCTTCACTCTGTGAATATCACCTGGGTGCTGAGAA 911
QY     254 pGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGluValVally 274
Db     912 TGAATTTAAACATGCAAGTTCGCAATTTCTCGGCGGTAGTGTGGGCTTTTCAGGTGGTGA 971
QY     274 sGlyPheAsnLeuTyLysLysGlyValTySerSerAspThrCysGlyArgAspProMe 294
Db     972 TGGTTTCCATTTACAGAAATGGAGTTTACACTAGTACATTTGGGTAGCATTCCCA 1031
QY     294 tAspValAsnHisAlaValLeuAlaValGlyTyThrGlyValGluAspGlyIleProTyTr 314
Db    1032 GGATGTGAACCATGCGCTGCTGCTGCGGTATGGAGTTGAGAAATGGCGTCCCATATTG 1091
QY     314 pleuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyThrPheLysMetGluLe 334
Db    1092 GCTCATAAAATTCATGGGAGAAAGTTGGGTGAGATGGCTACTTCAAAATGGAAT 1151
QY     334 uGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyProIleValAla 351
Db    1152 GGGGAAGAACATGTGTGGTGTTCGAACCTTGTGGCTTATCCAGTTGTGGCA 1203

```

RESULT 14

```

US-10-219-220-241
; Sequence 241, Application US/10219220
; Publication No. US20030082724A1
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; FILE OF INVENTION: death and their use in the modification of plant development
; FILE REFERENCE: 11000.102261
; CURRENT APPLICATION NUMBER: US/10/219, 220
; PRIOR FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: U.S. No. US20030082724A1 09/325,932
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 290
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 241
; LENGTH: 1652

```

```

; TYPE: DNA
; ORGANISM: Pinus radiata
US-10-219-220-241
Alignment Scores:      Length:      1652
Pred. No.:            2.3e-140      Matches:      228
Score:                1148.00      Conservative: 42
Percent Similarity:   75.21%      Mismatches:   79
Best Local Similarity: 63.51%      Indels:       10
Query Match:          61.55%      Gaps:         8
DB:                   14

US-10-087-714-2 (1-352) x US-10-219-220-241 (1-1652)
QY      1  MetAlaAlaLysLeuLeuPheLeu-----LeuPheLeuValSerAlaLeuSerVal 18
Db     264 ATGCTCGCGTCTCGCCATTGTTCTGAGCACATTTGTTTGTCTG---GCGTATCTGTA 320
QY     19  AlaLeuAlaGly---PheGluGluAspAsnProIleArgSerValThrGlnArgProAsp 37
Db     321 TCTGCTCGCGGTGCTTTGAAGAACTGGGTATATTGATATGTCCTCACTGCAAGATCCAA 380
QY     38  SerIleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAla 57
Db     381 GACCTGGAGTCCAAATTCGTTAAATCTTGGGAACCAATCCCAAGTCTCTACAAATTGCG 440
QY     58  ArgPheAlaArgArgTyGlyLysSerTyGlySerGluGluGluIleLysLysArgPhe 77
Db     441 GAGTTCTGCTGCAGATATGCAAGAGGTACGATTTCTGTCATCAGCTTGTGCATAGATT 500
QY     78  GlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTy 97
Db     501 AATCTCTTTGTGAAGAACGTGGAGCTGATCGAGTCAAGAACAGAAATGCAAGTTCCTTAT 560
QY     98  ThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeu 117
Db     561 ACTTTGCAATAATGAGTTTGCTGACATAACATGCGAGGAATTCATGACAAATATTG 620
QY    118  GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyVal 137
Db     621 GGTCTTTACAGAACTGTTCTGGCTACC---CACAGTAAACCATAGTTGAGTATGCCAG 677
QY    138  LeuProValThrArgAspTrpArgGluGlnGlyIleValSerProValIleAspGlnGly 157
Db     678 CTTCCTCGGAAGAAAGACTGGAGACAGAGGCACTAGTGTCTGTATAAAACCAAGCC 737
QY    158  SerCysGlySer---TrpThrPheSerThrGlyAlaLeuGluAlaAlaTyThrGln 176
Db     738 CATTTGGATCTCTGCTGACATTCAGCACTACTGGAGCACTAGAGGCTGCTATACTCAG 797
QY    177  LeuThrGlySerThr-----LeuSerGluGlnGlnLeuValAspCysAlaSerAlaPhe 194
Db     798 GCTACAGAAAGACTGTTATCTCTGTAACAGCAGCTGTTGACTGTGCTGGAGCATTT 857
QY    195  AsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyValLysTyAsn 213
Db     858 AACAACTTTGTTGCAATGGTGGACTGCCATCCCAAGCAATTTGAGTATATCAAGTACAA 917
QY    214  GlyGlyIleAspThrGluGlnThrTyProTyThrLeuGlyValMetGlyIleCysAsnPhe 233
Db     918 GGAGGCTTGAACACTGAGGAAGCCTATCTCTTACACGGCTAAAGATGGTGTGTTGAAATAT 977
QY    234  LysGlnGluAsnValGlyValIleAspSerIleAsnIleThrLeuGlyAlaGlu 253
Db     978 GATGTAAACATGTCGGTGTAAAGGTTGCTGATAGTGTCAACATCAGTTTGGGGGAGAA 1037
QY    254  AspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGluValVal 273
Db    1038 GATGACCTCAAAGCTGCTGGGTCTAGTTGCGCCTGTGCTAGTGGCATTCCAAGTAAAT 1097
QY    274  LysGlyPheAsnLeuTyLysLysGlyValTySerSerAspThrCysGlyArgAspPro 293
Db    1098 TCGGAATTTTCGATTTTACAGAGAGAGTCTTTACAGCAACAGTGTGTGTCAGAGGCCA 1157

```

```
Qy 294 MetAspValAsnHisAlaValLeuAlaValGlyTyrClyVal---GluAspGlyIlePro 312
Db 1158 ATGATGTCACCATGCTGTTTGGCGTTGGTATGGTTAGTGTAGGAGGGGACTCCA 1217
Qy 313 TyrTrpLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLysMet 332
Db 1218 CACTGGATCATCAAGATTCTCTGGGAAAGAGCTGGGGTCTTGATGATACACTTCAAGATG 1277
Qy 333 GluLeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIleValAla 351
Db 1278 GAGTTAGGGAAGAAATATGTGTGGTGTGGCACTTGGCTTCGTACCCATTGTTTCT 1334

RESULT 15
US-10-425-114-3201
; Sequence 3201, Application US/10425114
; Publication No. US20040034889A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425.114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 3201
; LENGTH: 830
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700243475_FLI
US-10-425-114-3201

Alignment Scores:
Pred. No.: 6,39e-102 Length: 830
Score: 852.50 Matches: 158
Percent Similarity: 89.06% Conservative: 13
Best Local Similarity: 82.29% Mismatches: 18
Query Match: 45.71% Indels: 3
DB: 12 Gaps: 2

US-10-087-714-2 (1-352) x US-10-425-114-3201 (1-830)
Qy 163 ThrPheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGlySer----- 180
Db 4 ACCCTTCAGCACTACTGGTGCACTTGGAGCTGCATATACCAGGCAACTGGCAAGCCCATC 63
Qy 181 ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPheGlyCys--- 199
Db 64 TCTCTCTGAGCAACAGACTTATTGACTGTGTTTGGATTCATCAAAATTTGGATGCAAC 123
Qy 200 GlyGlyLeuProSerGlnAlaPheGluTyrVallyTyrAsnGlyGlyIleAspThrGlu 219
Db 124 GGAGGCTTCCTCCATCCAGGCTTGGATACATCAATCAATGTTGGCTTGACACTGAG 183
Qy 220 GlnThrTyrProTyrLeuGlyValMetGlyIleCysAsnPheLysGlnGluAsnValGly 239
Db 184 GAATCTTACCCCTTACCAGGCTGTCAATGGAATCTGCAAGTTTAAAGATGAGAATGTTGA 243
Qy 240 ValLeuValIleAspSerIleAsnIleThrLeuGlyAlaGluAspGluLeuLysHisAla 259
Db 244 GTCAGGTTTGGACTTCGGTTACATCACCTGGGTGGCTGAGGATGAACTGAAGGATGCT 303
Qy 260 ValGlyLeuValArgProValSerValAlaPheGluValVallyGlyPheAsnLeuTyr 279
Db 304 GTTGGTCTGGTTCGCCAGTAGTGTGCTTCGAGGTGATCACTGGTTTCAGGCTGTAC 363
Qy 280 LysLysGlyVallyTyrSerSerAspThrCysGlyArgAspProMetAspValAsnHisAla 299
```

```
Db 364 AAGAGCGGAGTTTACCTAGCGACCATTTGTGAACTACACCGATGGATGTGAACACGCT 423
Qy 300 ValLeuAlaValGlyTyrGlyValGluAspGlyIleProTyrTrpLeuIleLysAsnSer 319
Db 424 GTTCTGGGCTGTTGGCTACGGGTGCGAAGATGGTGTACCTACTGGCTCATCAAGAACTCA 483
Qy 320 TrpGlyThrAsnTrpGlyAspAsnGlyTyrPheLysMetGluLeuGlyLysAsnMetCys 339
Db 484 TGGGCGCTGACTGGGGTGTATGAGGGTTACTTCAAGATGGAATGGGCAAGACATGTGC 543
Qy 340 GlyValAlaThrCysAlaSerTyrProIleValAla 351
Db 544 GGTGTGTGTACGTGTGCATCCTACCTTATTGTGCA 579
```

Search completed: March 30, 2004, 01:10:56
Job time : 342 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 29, 2004, 22:03:58 ; Search time 33 Seconds
(without alignments)
1026.043 Million cell updates/sec

Title: US-10-087-714-2

Perfect score: 1865

Sequence: 1 MAAKLLFFLLFLVSALSVAL.....ELGKNMCGVATCASYPIVAV 352

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.*

1: Pirl.*

2: Pirl.*

3: Pirl.*

4: Pirl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1325.5	70.9	360	2	S59598
2	1281.5	68.7	362	1	KHRZOG
3	1254.5	67.3	350	2	S71923
4	1252	67.1	361	1	KHBH
5	1249	67.0	356	2	S56348
6	1207.5	64.7	377	2	T47471
7	796	42.7	333	1	KERTH
8	763	40.9	335	1	KHHUH
9	655.5	35.2	326	2	S43991
10	616	33.0	374	2	T03941
11	613.5	32.9	338	2	JCS443
12	609	32.7	462	2	JN0719
13	608	32.6	364	2	T12039
14	604.5	32.4	326	2	S53027
15	604	32.4	364	2	T46630
16	603.5	32.4	339	2	A53810
17	594.5	31.9	323	2	S19650
18	594.5	31.9	368	2	S47312
19	587	31.5	471	1	KHRZOB
20	586	31.4	322	2	S19649
21	585.5	31.4	455	2	T12041
22	584.5	31.3	344	2	JX0366
23	577.5	31.0	466	2	T06416
24	573	30.7	324	2	S47432
25	573	30.7	376	2	E85435
26	572.5	30.7	454	2	JC4848
27	571.5	30.6	328	2	T09259
28	570	30.6	326	2	JQ1121
29	566.5	30.4	331	2	JCS442

RESULT 1
S59598
cysteine proteinase (EC 3.4.22.-) 2 precursor [similarity] - maize
C/Species: Zea mays (maize)
C/Date: 15-Feb-1996 #Sequence_revision 01-Mar-1996 #text_change 20-Jun-2000
C/Accession: S59598
R/Domoto, C.; Watanabe, H.; Abe, M.; Abe, K.; Arai, S.
Biochim. Biophys. Acta 1263, 241-244, 1995
A/Title: Isolation and characterization of two distinct cDNA clones encoding corn seed
A/Reference number: S59597; MUID:96004895; PMID:7548211
A/Accession: S59598
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-360 <DOM>
A/Cross-references: EMBL:D45403; NID:G1688044; PIDN:BAA08245.1; PID:9644490
C/Superfamily: papain
C/Keywords: cysteine proteinase; hydrolase
F:167,307,321/Active site: Cys, His, Asn #status predicted

Query Match 70.9%; Score 1322.5; DB 2; Length 360;
Best Local Similarity 71.7%; Pred. No. 5.8e-103;
Matches 258; Conservative 28; Mismatches 65; Indels 9; Gaps 6;

QY	1	MAAKLLFFLLFLVSALSVAL--AGEEDNPISRTORPDS--TEPALLGLVGSCHAFHFA	57
DB	1	MPRRFLVLAVVLADTAAVNSGFADSPIRPVTDRASALESTVFAALGTRDALRFA	60
QY	58	RPARRYGKSYGSEEEIKKFGIFVENLAFIRSTNRKLSYTLGINQFADLTWEEFTNRL	117
DB	61	RAVRVYKSYESAEEVHKFRIFSESLSQLVRSNKKLSYRLGINRFDMSWEFPRATRL	120
QY	118	GAQONCSATAHGNHRP--VDGVLPTVRDREOGIVSPYKQDQSCGS-WTFSTTGALAAAY	174
DB	121	GAQONCSATLTGNHRRMAAVALPETKDRWDGIVSPVKNQHGCGSCWTFSTTGALAAAY	180
QY	175	TOLTGS--TLSEQLVDCAFAFNFC--GGLPSQAFYVYKNGGIDTEQTYPLGYMGIC	231
DB	181	TQATGFISSLSSQQLVDCGFANFNFCNGGLPSQAFYIKYNGGLDTEESYPQGVNGIC	240
QY	232	NFKQENGVKVIDSINITGAEDELKHVGLVRPVSVAEVKGNLYKKGIVSSFTGCR	291
DB	241	KFKENGVKVLDSVNIITLGADELKDAVGLVRPVSVAFEVITGTFRLYKSGVYTSHCCT	300
QY	292	DPMDVNHAVLAVGYGVEDGIPYWLKNSMGTNWNGYFKMELGKNMCGVATCASYPIVA	351
DB	301	TPMDVNHAVLAVGYGVEDGVPYWLKNSMGADNWDGEGYFKMENGKNMCGVATCASYPIVA	360

RESULT 2
KHRZOG
oryzain (EC 3.4.22.-) gamma precursor - rice
C/Species: Oryza sativa (rice)

30	566.5	30.4	362	2	T09528	probable cysteine
31	566	30.3	325	2	S49451	cysteine proteinas
32	563.5	30.2	363	2	S11862	cysteine proteinas
33	562.5	30.2	331	2	JCS441	cathepsin L-like c
34	560.5	30.1	378	2	S47434	cysteine proteinas
35	559.5	30.0	320	2	S19651	cysteine proteinas
36	559.5	30.0	363	2	T06726	actinidin (EC 3.4
37	559	30.0	380	1	TAGB	cysteine proteinas
38	558	29.9	358	2	S42882	prestalk cathepsin
39	558	29.9	376	1	KHDOP	cathepsin L (EC 3.
40	557.5	29.9	313	2	S47433	cathepsin L (EC 3.
41	557	29.9	317	2	S44151	cysteine proteinas
42	556.5	29.8	480	2	T01207	cysteine proteinas
43	555.5	29.8	427	2	S57776	oryzain (EC 3.4.22
44	554.5	29.7	458	1	KHRZOA	probable cysteine
45	554	29.7	337	2	T24387	

ALIGNMENTS

C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jun-2000
C:Accession: J00390; C40053
R;Watanabe, H.; Abe, K.; Emori, Y.; Hosoyama, H.; Arai, S.
submitted to JIPID, May 1991
A:Reference number: J00388
A:Accession: J00390
A:Molecule type: mRNA
A:Residues: 1-362 <WAT1>
R;Watanabe, H.; Abe, K.; Emori, Y.; Hosoyama, H.; Arai, S.
J. Biol. Chem. 266, 16897-16902, 1991
A:Title: Molecular cloning and gibberellin-induced expression of multiple cysteine protease
A:Reference number: A40053; MUID:91358494; PMID:1885617
A:Accession: C40053
A:Molecule type: mRNA
A:Residues: 1-362 <WAT2>
A:Cross-references: GB:D90408; NID:G218184; PIDN:BAAL4404.1; PID:G218185
C:Superfamily: papain
C:Keywords: cysteine proteinase; glycoprotein; hydrolase; seed
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-144/Domain: propeptide #status predicted <PRO>
F;145-362/Product: oryzain gamma #status predicted <MAT>
F;128,258/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;166-209,200-242,300-350/Disulfide bonds: #status predicted
F;169,309,329/Active site: Cys, His, Asn #status predicted
Query Match 68.7%; Score 1281.5; DB 1; Length 362;
Best Local Similarity 66.9%; Pred. No. 1.6e-99;
Matches 242; Conservative 45; Mismatches 64; Indels 11; Gaps 6;
QY 1 MAAKLLFFLLFVLVLAAGFEEDNPIRSTQRPDS-IEPAILGLVGSORHAF 54
DB 1 MAARRILLLAAVAATSAVAASGDFDSDNPIRSTVTHAASALESTVIAALGRTRGAL 60
QY 55 HFARFARYKSGSSEIEIKRFGIFVENIAFIRSTRNRKDLSTYLGINOFADLTWEEFT 114
DB 61 RFARFARVHGKRGDAEVRQRRIFSESLELVKSTNRGLPYRLGINFADMSWEEFOA 120
QY 115 NRGAAQNCATAGHNRHFDVG-VLPVTRDRWQGVSPVKQSGCS-WTFSTTGALRA 172
DB 121 SRGAAQNCATAGHNRHFDALPETKDRDGGIVSPVKQSGCSWPFSTTGLSLA 180
QY 173 AYTQLTGS--TLSEQQVLDCASAFNNFGC-GGLPSQAFYVKYNGIDTEQTYPYLVGNG 229
DB 181 RYTOATGPPVSLSEQQVLDCATRYNNFGCSGGLPSQAFYIKYNGIDTEETAYPYTVNG 240
QY 230 ICNFKQENVGVKVIDSNITLGADELKHAVALRVPVSAFVVKGNLYKGVYSSDTC 289
DB 241 ICHYPENAGVKVLDVSNITLVADELKHAVALRVPVSAFVVKGNLYKGVYSSDTC 300
QY 290 GRDPMVNHAVLAGYGVEDGIPYWLTKNSWGTNWDGNGYFMELGKNCMGVATCASYPI 349
DB 301 GTSPMDVNHAVLAGYGVNGVPYWLTKNSWGTNWDGNGYFMELGKNCMGVATCASYPI 360
QY 350 VA 351
DB 361 VA 362
RESULT 3
S71923
Cysteine proteinase (EC 3.4.22.-) - garden pea
C:Species: Pisum sativum (garden pea)
C>Date: 04-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 20-Jun-2000
C:Accession: S71923
R;Jones, C.G.; Tucker, G.A.; Lycett, G.W.
Biochim. Biophys. Acta 1296, 13-15, 1996
A:Title: Pattern of expression and characteristics of a cysteine proteinase cDNA from ge
A:Reference number: S71923; MUID:96350414; PMID:8765223
A:Accession: S71923
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-350 <JON>
A:Cross-references: EMBL:Z68291; NID:G1134881; PIDN:CAA92583.1; PID:G1134882

C:Superfamily: papain
C:Keywords: cysteine proteinase; hydrolase
F;157,297,317/Active site: Cys, His, Asn #status predicted
Query Match 67.3%; Score 1254.5; DB 2; Length 350;
Best Local Similarity 67.2%; Pred. No. 2.7e-97;
Matches 240; Conservative 39; Mismatches 65; Indels 13; Gaps 6;
QY 1 MAAKLLFFLLFVLVLAAGFEEDNPIRSTQRPDS-IEPAILGLVGSORHAF 58
DB 1 MAQWSLLLVLCFVAS---AAAGSFHSDNPIRMVS---DVEBQLLQVIGESHAVSFAR 53
QY 59 FARRYKSGSSEIEIKRFGIFVENIAFIRSTRNRKDLSTYLGINOFADLTWEEFTNRLG 118
DB 54 FARRYKSGSSEIEIKRFGIFVENIAFIRSTRNRKDLSTYLGINOFADLTWEEFTNRLG 113
QY 119 AAQNCSTAHGNHFRFVDGVLVTRDWRWQGVSPVKQSGCS-WTFSTTGALRAAYTQL 177
DB 114 AAQNCSTAHGNHFRFVDGVLVTRDWRWQGVSPVKQSGCS-WTFSTTGALRAAYTQL 173
QY 178 TGS--TLSEQQVLDCASAFNNFGC-GGLPSQAFYVKYNGIDTEQTYPYLVGNGICNFK 234
DB 174 FGKNSLSEQQVLDCASAFNNFGC-GGLPSQAFYVKYNGIDTEQTYPYLVGNGICNFK 233
QY 235 QENVGVKVIDSNITLGADELKHAVALRVPVSAFVVKGNLYKGVYSSDTCGRDPM 294
DB 234 SEHVAVKVLGSVNITLGADELKHAVALRVPVSAFVVKGNLYKGVYSSDTCGRDPM 293
QY 295 DVNHAVLAGYGVEDGIPYWLTKNSWGTNWDGNGYFMELGKNCMGVATCASYPIVA 351
DB 294 DVNHAVLAGYGVEDGIPYWLTKNSWGTNWDGNGYFMELGKNCMGVATCASYPIVA 350
RESULT 4
KHEH
Aleurain (EC 3.4.22.-) precursor - barley
C:Species: Hordeum vulgare (barley)
C>Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 04-Oct-1996
C:Accession: A25492; B25492
R;Rogers, J.C.; Dean, D.; Heck, G.R.
Proc. Natl. Acad. Sci. U.S.A. 82, 6512-6516, 1985
A:Title: Aleurain: a barley thiol protease closely related to mammalian cathepsin H.
A:Reference number: A25492; MUID:86016732; PMID:3901004
A:Accession: A25492
A:Molecule type: DNA
A:Residues: 1-361 <ROG1>
A:Residues: B25492
A:Molecule type: mRNA
A:Residues: 1-361 <ROG2>
A:Cross-references: GB:X05167; GB:M11760; GB:M15209; GB:Y00107
C:Comment: Aleurain is synthesized by the aleurone cells stimulated by gibberellic or
C:Superfamily: papain
C:Keywords: aleurone cell; cysteine proteinase; germination; glycoprotein; hydrolase;
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-143/Domain: propeptide #status predicted <PRO>
F;144-361/Product: aleurain #status predicted <MAT>
F;165-208,199-240,298-348/Disulfide bonds: #status predicted
F;168,307,327/Active site: Cys, His, Asn #status predicted
F;188,256/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 67.1%; Score 1252; DB 1; Length 361;
Best Local Similarity 67.3%; Pred. No. 4.6e-97;
Matches 241; Conservative 42; Mismatches 65; Indels 10; Gaps 7;
QY 3 AKLLFFLLFVLVLAAGFEEDNPIRSTQRPDS-IEPAILGLVGSORHAF 58
DB 4 ARVLLALLAVLATAVAVASSSFDSPNPIRPTDRAASTLESVAVLGALGRTRHAF 63
QY 59 FARRYKSGSSEIEIKRFGIFVENIAFIRSTRNRKDLSTYLGINOFADLTWEEFTNRLG 118
DB 64 FARRYKSGSSEIEIKRFGIFVENIAFIRSTRNRKDLSTYLGINOFADLTWEEFTNRLG 123
QY 119 AAQNCSTAHGNHFRFVDGVLVTRDWRWQGVSPVKQSGCS-WTFSTTGALRAAYTQ 176

Db 124 AAQTCATLGNHLMRDAALPETKDWREDDGIVSPVKNOAHGCGSCWTFSTTGALEAAATQ 183
QY 177 LTGS--TLSEQQLVDCASAFNNFGC--GGLPSQAFYVYKNGGIDTEQTYPLGVGMGICNF 233
Db 184 ATGNKISLSEQQLVDCAGFNFGCNGGLPSQAFY--QYNGGIDTESYPYKGVNGVCHY 242
QY 234 KOENVGVKVIDSINITLGADELKHAVALVRPVSVAFAEVVKGFLNLYKGVYSSDTCGRDP 293
Db 243 KAENAAVQLVDSVNTLNAEDLNKNAVLVRPVSVAFAEQVIDGFRQYKGVYSSDTCGRDP 302
QY 294 MDVNHAVLAGYGVDEGIPYLLIKNSWGTNGDNGYFKMELGKMGVATCASYPIVA 351
Db 303 DDVNHAVLAGYGVDEGIPYLLIKNSWGTNGDNGYFKMELGKMGVATCASYPIVA 360

RESULT 5
S66348
cysteine proteinase (EC 3.4.22.-) senescence-associated precursor (clone SEN3) [similar
C;Species: Lycopersicon esculentum (tomato)
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
C;Accession: S66348
R;Drake, R.; John, I.; Farrell, A.; Cooper, W.; Schuch, W.; Grierson, D.
Plant Mol. Biol. 30, 755-767, 1996
A;Title: Isolation and analysis of cDNAs encoding tomato cysteine proteases expressed du
A;Reference number: S66348; MUID:96194458; PMID:8624407
A;Accession: S66348
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-356 <DRA>
A;Cross-references: GB:248736; NID:g1235544; PIDN:CAA08629.1; PID:g1235545
C;Superfamily: papain
C;Keywords: cysteine proteinase, hydrolase
F;163,303,323/Active site: Cys, His, Asn #status predicted

Query Match 67.0%; Score 1249; DB 2; Length 356;
Best Local Similarity 68.3%; Pred. No. 8.1e-97;
Matches 243; Conservative 35; Mismatches 70; Indels 8; Gaps 5;

QY 3 AKLLFFLLFLVSALSVALAG---FEEDNPIRSVTQRPDSIEPAILGVLSGCRHAFHARF 59
Db 2 SRLSLVILVAGLFATALAGPATFADKNPIQVY--FPDELENGILQVGVQTRALSARF 60
QY 60 ARRYKSGSSEETIKKRFIVENLAFIRSTRNKLDSYTLGINOFADLTWEEFRTNRLG 119
Db 61 AIRHKRVDSVEELKQRFIEFLDNLMKLRSHNRKGLAYKLGINEFTDLTWDEFKHLGA 120
QY 120 AONCSATAHGNHRFVDGVLVTRWRQGVIVSPVKDQSCGS--WTFSTTGALEAAATQ 178
Db 121 SONGSATTGKLNLTNVVLPETKDWKRDGIVSPVKAQCKGCGSCWTFSTTGALEAAATQ 180
QY 179 GS--TLSEQQLVDCASAFNNFGC--GGLPSQAFYVYKNGGIDTEQTYPLGVGMGICNFQ 235
Db 181 KGISLSEQQLVDCAGFNFGCNGGLPSQAFYVYKNGGLDTEAEYPTGKGVCKFSQ 240
QY 236 ENVGKVIDSINITLGADELKHAVALVRPVSVAFAEVVKGFLNLYKGVYSSDTCGRDPMD 295
Db 241 ANIGVKVIVSSVNTILGAELKHAVALVRPVSVAFAEVVKGFLNLYKGVYSSDTCGRDPMD 300
QY 296 VNHAVLAGYGVDEGIPYLLIKNSWGTNGDNGYFKMELGKMGVATCASYPIVA 351
Db 301 VNHAVLAGYGVDEGIPYLLIKNSWGTNGDNGYFKMELGKMGVATCASYPIVA 356

RESULT 6
T47471
cysteine proteinase (EC 3.4.22.-) F18N11.70 [similarity] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 26-May-2000
C;Accession: T47471
R;Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unsel, M.; Mewes, H.W.; Lemcke, K.
submitted to the Protein Sequence Database, February 2000
A;Reference number: 224467

A;Accession: T47471
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-377 <DOR>
A;Cross-references: EMBL:AL132953
A;Experimental source: cultivar Columbia; BAC clone F18N11
C;Genetics:
A;Map position: 3
A;Introns: 65/2; 106/1; 144/3; 169/2; 257/3; 301/3; 345/2
A;Note: F18N11.70
C;Superfamily: papain
C;Keywords: cysteine proteinase, hydrolase
F;165,305,325/Active site: Cys, His, Asn #status predicted

Query Match 64.7%; Score 1207.5; DB 2; Length 377;
Best Local Similarity 65.9%; Pred. No. 2.6e-93;
Matches 228; Conservative 37; Mismatches 74; Indels 7; Gaps 4;

QY 1 MAAKLIF---FLLFLVSALSVALAGFEEDNPIRSVTQRPDSIEPAILGVLSGCRHAFHFA 57
Db 1 MSVKLNLSSSIILLIFAAASKEIGFDESNPDKMVDNLHELEDITVVQILGOSRHVLSFS 60
QY 58 RPARRYKSGSSEETIKKRFIVENLAFIRSTRNKLDSYTLGINOFADLTWEEFRTNRL 117
Db 61 RTHRGKGYQSVVEEMKLFSPVFKENLDLIRSTNKKGLSYKLSLQAFADLTWEEFQRYKL 120
QY 118 GAAQNCSTAHGNHRFVDGVLVTRWRQGVIVSPVKDQSCGS--WTFSTTGALEAAATQ 176
Db 121 GAAQNCSTAHGNHRFVDGVLVTRWRQGVIVSPVKDQSCGS--WTFSTTGALEAAATQ 180
QY 177 LTGS--TLSEQQLVDCASAFNNFGC--GGLPSQAFYVYKNGGIDTEQTYPLGVGMGICNF 233
Db 181 AFGKISLSEQQLVDCAGFNFGCNGGLPSQAFYVYKNGGLDTEAEYPTGKGVCKF 240
QY 234 KOENVGVKVIDSINITLGADELKHAVALVRPVSVAFAEVVKGFLNLYKGVYSSDTCGRDP 293
Db 241 SAKNIGVQRDSVNTILGADELKHAVALVRPVSVAFAEVVKGFLNLYKGVYSSDTCGRDP 300
QY 294 MDVNHAVLAGYGVDEGIPYLLIKNSWGTNGDNGYFKMELGKMGVATCASYPIVA 351
Db 301 MDVNHAVLAGYGVDEGIPYLLIKNSWGTNGDNGYFKMELGKMGVATCASYPIVA 356

RESULT 7
KHRTH
cathepsin H (EC 3.4.22.16) precursor - rat
N;Alternate names: cathepsin B3; cathepsin Ba
C;Species: Rattus norvegicus (Norway rat)
C;Date: 18-Apr-1984 #sequence_revision 30-Jun-1993 #text_change 21-Jul-2000
C;Accession: S00211; S05213; A00976; A60371
R;Ishidoh, K.; Imaizoh, S.; Emori, Y.; Ohno, S.; Kawasaki, H.; Kominami, E.;
FEBS Lett. 226, 33-37, 1987
A;Title: Molecular cloning and sequencing of cDNA for rat cathepsin H. pro-
A;Reference number: S00211; MUID:88083632; PMID:3691815
A;Accession: S00211
A;Molecule type: mRNA
A;Residues: 1-333 <ISH>
A;Cross-references: GB:M36320; NID:g55885; PIDN:CAA68699.1; PID:g55886
R;Ishidoh, K.; Kominami, E.; Katunuma, N.; Suzuki, K.
FEBS Lett. 253, 103-107, 1989
A;Title: Gene structure of rat cathepsin H.
A;Reference number: S05213; MUID:89338722; PMID:2759235
A;Accession: S05213
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-286, 'A', 288-333 <ISH2>
R;Takio, K.; Towatari, T.; Katunuma, N.; Teller, D.C.; Titani, K.
Proc. Natl. Acad. Sci. U.S.A. 80, 3666-3670, 1983
A;Title: Homology of amino acid sequences of rat liver cathepsins B and H with that of
A;Reference number: A00976; MUID:83221657; PMID:6574504
A;Accession: A00976
A;Molecule type: protein
A;Residues: 114-333 <TAK>

R;Qian, F.; Frankfater, A.; Miller, R.V.; Chan, S.J.; Steiner, D.P.
 Int. J. Biochem. 22, 1457-1464, 1990
 A;Title: Molecular cloning of rat precursor cathepsin H and the expression of five lysos
 A;Reference number: A60371; MUID:91114797; PMID:2276418
 A;Accession: A60371
 A;Molecule type: mRNA
 A;Residues: 4-333 <Q1A>
 A;Cross-references: GB:M38135
 C;Comment: Cathepsin H is composed of a minichain and a large chain. The large chain may
 C;Genetics:
 A;Introns: 29/1; 39/3; 75/1; 98/3; 133/3; 162/3; 181/2; 208/3; 231/3; 267/2; 309/2
 C;Superfamily: papain
 C;Keywords: cysteine proteinase; glycoprotein; hydrolase; lysosome
 F;1-21/Domain: signal sequence #status predicted <PRO>
 F;96-102/Product: cathepsin H minichain #status predicted <MCH>
 F;114-333/Product: cathepsin H large chain #status experimental <MCH>
 F;114-290/Product: cathepsin H heavy chain #status predicted <HCH>
 F;291-333/Product: cathepsin H light chain #status predicted <LCH>
 F;70-99/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;136-179-170-212-270-320/Disulfide bonds: #status predicted
 F;139-279-299/Active site: Cys, His, Asn #status predicted
 F;228/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 42.7%; Score 796; DB 1; Length 333;
 Best Local Similarity 52.1%; Pred. No. 6.2e-59;
 Matches 159; Conservative 50; Mismatches 88; Indels 8; Gaps 7;

QY 54 FHPARFARYKSGSEBEIKRFGIFVENLAFIRSTNEKDLSTYLGINQPADLTWEERF 113
 DB 31 FHTSMWKQHKTYT-SREYSHRLQVFANNWKIQAHHNORHTFFMGLNPFSDMFPAIK 89

QY 114 TNRL-GAAQNCATAGHNRFPDGLVLPVTRDREQG-IVSPVKDQSCGS-WTFSTTGA 170
 DB 90 HKYLWSEPNQCSAT-KSNYLRLGTGTPPSSMDWRKKGNVSPVKNOGACGSCWTFSTTGA 148

QY 171 EAYTQTLTGS--TLSEQQLVDCASAFNNFGC-GGLPSQAFYKNGGIDTEQTPYLG 227
 DB 149 ESAVAIATGKMLSLAEQQLVDCAQDNFNNGCCQGLPSQAFYILNKGIMGDSYPIQK 208

QY 228 MGICNFQKQENVGVKVIDSINITLGADELKHADELGVLPVSVAFVVKGNLYKKGYS 287
 DB 209 NGQCKNPEKAVAFKVNVTITLDEAAVVEAVALNVPVSFAFEVTEDFMVKSGYSSN 268

QY 288 TCGRDPMDVNHAVLAVGYGVEDGIPYWLKNSWGTWNGDGFKNELGKMGCVATCASY 347
 DB 269 SCHKTPDKVNHAVLAVGYGEQNGLLYIVKNSWGNNGNYFLIERGKMGCLAACASY 328

QY 348 PIVAV 352
 DB 329 PIPQV 333

RESULT 8
 KHVH
 cathepsin H (EC 3.4.22.16) precursor [validated] - human
 C;Species: Homo sapiens (man)
 C;Date: 30-Sep-1990 #sequence revision 30-Jun-1991 #text change 08-Dec-2000
 C;Accession: S12486; S07634; S00818; S00322; A27011; C27011; S00635
 R;Fuchs, R.
 submitted to the EMBL Data Library, October 1989
 A;Reference number: S12486
 A;Accession: S12486
 A;Molecule type: mRNA
 A;Residues: 1-335 <FUC>
 A;Cross-references: EMBL:X16832; NID:929709; PIDN:CAA34734.1; PID:929710
 Nucleic Acids Res. 17, 9471, 1989
 A;Title: Nucleotide sequence of human preprocathepsin H, a lysosomal cysteine proteinase
 A;Reference number: S07634; MUID:90067944; PMID:2587265
 A;Accession: S07634
 A;Molecule type: mRNA
 A;Residues: 1-115 <FU2>

A;Cross-references: EMBL:X16832
 R;Fuchs, R.; Machleidt, W.; Gassen, H.G.
 Biol. Chem. Hoppe-Seyler 369, 469-475, 1988
 A;Title: Molecular cloning and sequencing of a cDNA coding for mature human kidney cat
 A;Reference number: S00818; MUID:89076480; PMID:2849458
 A;Accession: S00818
 A;Molecule type: mRNA
 A;Residues: 88-335 <FU3>
 A;Cross-references: EMBL:X07549; NID:929707; PIDN:CAA30428.1; PID:929708
 R;Ritonja, A.; Popovic, T.; Kotnik, M.; Machleidt, W.; Turk, V.
 FEBS Lett. 228, 341-345, 1988
 A;Title: Amino acid sequences of the human kidney cathepsins H and L.
 A;Reference number: S00322; MUID:88137635; PMID:3342889
 A;Accession: S00322
 A;Molecule type: protein
 A;Residues: 98-105;114-178 'H', 180-305 'E', 307-335 <RIT>
 R;Machleidt, W.; Ritonja, A.; Popovic, T.; Kotnik, M.; Brzin, J.; Turk, V.; Machleidt,
 in Cysteine Proteinases and Their Inhibitors, Turk, V., ed., pp.3-18, Walter de Gruyter
 A;Title: Human cathepsins B, H and L: characterization by amino acid sequences and som
 A;Reference number: A27011
 A;Contents: annotation
 C;Comment: Cathepsin H is composed of a minichain and a large chain. The large chain m
 C;Genetics:
 A;Gene: GDB:CTSH
 A;Cross-references: GDB:120602; OMIM:116820
 A;Map position: 15q24-15q25
 C;Superfamily: papain
 C;Keywords: cysteine proteinase; glycoprotein; hydrolase; lysosome
 F;1-22/Domain: signal sequence #status predicted <SIG>
 F;23-97/Domain: propeptide #status predicted <PRO>
 F;98-105/Product: cathepsin H minichain #status experimental <MCH>
 F;114-335/Product: cathepsin H large chain (minor form) #status experimental <MCH>
 F;116-335/Product: cathepsin H heavy chain #status experimental <MCH>
 F;116-232/Product: cathepsin H heavy chain #status experimental <HCH>
 F;293-335/Product: cathepsin H light chain #status experimental <LCH>
 F;72/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;101-230/Binding site: carbohydrate (Asn) (covalent) #status experimental
 F;102-327-338-181-172-214-272-322/Disulfide bonds: #status predicted
 F;141,281,301/Active site: Cys, His, Asn #status predicted

Query Match 40.9%; Score 763; DB 1; Length 335;
 Best Local Similarity 52.1%; Pred. No. 3.6e-56;
 Matches 159; Conservative 40; Mismatches 98; Indels 8; Gaps 7;

QY 54 FHPARFARYKSGSEBEIKRFGIFVENLAFIRSTNEKDLSTYLGINQPADLTWEERF 113
 DB 33 FHFWSMKHKRTY-STEYVHRLQTFASNRKINAHNNHHTFFMGLNPFSDMFPAIK 91

QY 114 TNRL-GAAQNCATAGHNRFPDGLVLPVTRDREQG-IVSPVKDQSCGS-WTFSTTGA 170
 DB 92 HKYLWSEPNQCSAT-KSNYLRLGTGTPPSSMDWRKKGNVSPVKNOGACGSCWTFSTTGA 150

QY 171 EAYTQTLTGS--EQQLVDCASAFNNFGC-GGLPSQAFYKNGGIDTEQTPYLG 227
 DB 151 ESAIATATGKMLSLAEQQLVDCAQDNFNNGCCQGLPSQAFYILNKGIMGDSYPIQK 210

QY 228 MGICNFQKQENVGVKVIDSINITLGADELKHADELGVLPVSVAFVVKGNLYKKGYS 287
 DB 211 DGYCKFQPKAIGFVKDVAITIVDEAMVEAVALNVPVSFAFEVTEDFMVKRTGIYSST 270

QY 288 TCGRDPMDVNHAVLAVGYGVEDGIPYWLKNSWGTWNGDGFKNELGKMGCVATCASY 347
 DB 271 SCHKTPDKVNHAVLAVGYGEKNGIPYIVKNSWGNNGNYFLIERGKMGCLAACASY 330

QY 348 PIVAV 352
 DB 331 PIPLV 335

RESULT 9
 S43991
 cathepsin L-like proteinases (EC 3.4.22.-) - liver fluke
 C;Species: Fasciola hepatica (liver fluke)

C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 04-Feb-2000
 C;Accession: S43991
 R;Wijffels, G.L.; Panaccio, M.; Salvatore, L.; Wilson, L.; Walker, I.D.; Spitchill, T.W. Biochem. J. 295, 781-790, 1994
 A;Title: The secreted cathepsin L-like proteinases of the trematode, Fasciola hepatica, A;Reference number: S43991; MUID:94250242; PMID:8192668
 A;Accession: S43991
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-326 <WJ>
 A;Cross-references: GB:L33771; NID:G497699; PIDN:AAA29136.1; PID:G497700
 A;Note: the sequence from Fig. 6 is inconsistent with that from Fig. 5 in having 57-Pro
 C;Superfamily: papain
 C;Keywords: cysteine proteinase; hydrolase
 F:132,269,289/Active site: Cys, His, Asn #status Predicted

Query Match 35.2%; Score 656.5; DB 2; Length 326;
 Best Local Similarity 44.6%; Pred. No. 2.8e-47;
 Matches 144; Conservative 52; Mismatches 102; Indels 25; Gaps 13;

Qy 44 LGVLGSCRHAFHAFARFYKSYGSEEEIKKRFIGFVENLAFIRSTN-RKDL---SYTL 99
 Db 11 VGVLSNDLWH--QWKRMYNKEYNGADDQHR-NIWEKNVKHIEHNLRLHDLGLVITYL 67

Qy 100 GINQFADLTWEEFRTNRLGAQNC-ATAHG-----NHRFVGVLPVTRDWRQGIQVSPV 153
 Db 68 GLNQFTDMTFEEFKAKYLTMSRASDILSHGVPEYANRAV---PKIDWRESGYVTEV 123

Qy 154 KDQSCGS-WTFSTTGALAAAY--TQLTGSTLSQQLVDCASAFNFGC-GGLPSQAF 209
 Db 124 KDQSCGSWAFSTITGWEGYMKNERTISFSQQLVDCSGPGNNGCGSLMENAYQY 183

Qy 210 VKYNGGIDTEQTYPYLGVMGICNFQKQNVGVKVIDSINITLGADELKHAVALVPRVSV 259
 Db 184 LK-QFGLTESSPYTAVEGQCRYNKQGVAKVTGYTVHSGSEVELKLVGARRPAVA 242

Qy 270 FEVKGENLYKGVYSSDTQGRDPMVNHAVLAGYGVDEGIPYWLKNSWGTNMGDNGY 329
 Db 243 VDVESDFMWRSGYIQSGTC--SPLRVNHAVLAGYGTGGTDTWIVNSWGTWYGERGY 300

Qy 330 FMELGK-NMCGVATCASYPTVA 351
 Db 301 IRVARNRGNMCGIASLASLPVA 323

RESULT 10
 T03941
 cysteine proteinase (EC 3.4.22.-) precursor - common tobacco
 C;Species: Nicotiana tabacum (common tobacco)
 C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
 C;Accession: T03941
 R;Becker, C.; Muentz, K.
 submitted to the EMBL Data Library, September 1997
 A;Description: CDNA cloning of a CPRI-homologous proteinase from germinating tobacco seed
 A;Reference number: Z15148
 A;Accession: T03941
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-374 <BC>
 A;Cross-references: EMBL:Z99173; PIDN:CAB16317.1
 A;Experimental source: clone TCPR1
 C;Superfamily: papain
 C;Keywords: cysteine proteinase; hydrolase
 F:1-20/Domain: signal sequence #status Predicted <SIG>
 F:21-374/Product: cysteine proteinase #status Predicted <MAT>
 F:164,299,319/Active site: Cys, His, Asn #status Predicted

Query Match 33.0%; Score 616; DB 2; Length 374;
 Best Local Similarity 40.4%; Pred. No. 8.3e-44;
 Matches 153; Conservative 58; Mismatches 112; Indels 56; Gaps 17;

Qy 1 MAKLIFLLF-LVSALVA---LAGFEEDNPIRSTVQRDPSIEPAILGVLSCHAFH 55

Db 1 MAKTIITLLFALPSSLGVAIDMSIIDYKNNHYARKWTLQSD--EDQVKV-----R 49

Qy 56 FARFARRYKSYGSEEEIKKRFIGFVENLAFIRSTNRK-DLSYTLGINQFADLTWEEPT 114

Db 50 YEWLAHGRNALNAGEKEKFEIPKDNLRFIEGHNNGRNYKYGLNQFADLTWEEPT 109

Qy 115 NELGAAQNCASATAGNHRFV-----DGLVPVTRDWRQGIQVSPVKDQSCGS 161

Db 110 MYLGTKSD-----ARRRFVSKSNPSQRYASRPNELMPSHSDVWRKRGAVAPIKNQSCGS 163

Qy 162 -WTFSTTGALAAAYTQLTGS--TLSEQQLVDCASAFNFGC-GGLPSQAFVVKYNGGID 217

Db 164 CWAFSTVAAVEGINQIVGEMITLSEQELVDC-DRVQNSGCGGLMDYAFEFIISNGMD 222

Qy 218 TEQTYPYLGVMGICNFQKQNVGVKVIDSINITLGADELKHAVALVPRVSVAFV 276

Db 223 TEKHPYRGVGEGRCDPVKRYKVKVSDIGYEDVPRNERALQKAVA-HQPVCAVAEASGRA 281

Qy 277 NLYKGVYSSDTQGRDPMVNHAVLAGYGVDEGIPYWLKNSWGTNMGDNGYFKWE--- 333

Db 282 QLYSGVFTGE-CGE---EVDHGVVWVGSGEDGVDYIVRNSWGTNMGDNGYVKNERNV 337

Qy 334 ----LQKNMCGVATCASYP 348

Db 338 KKSHLGK--CGIMTEASYP 354

RESULT 11
 JCS443
 cathepsin L-like cysteine proteinase (EC 3.4.22.-) cl [similarity] - Maize weevil
 C;Species: Sitophilus zeamais (maize weevil)
 C;Date: 17-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 20-Jun-2000
 C;Accession: JCS443
 R;Matsumoto, I.; Emori, Y.; Abe, K.; Arai, S.
 J. Biochem. 121, 464-476, 1997
 A;Title: Characterization of a gene family encoding cysteine proteinases of Sitophilus
 nd germ cells.
 A;Reference number: JCS441; MUID:97279038; PMID:9133615
 A;Accession: JCS443
 A;Molecule type: mRNA
 A;Residues: 1-338 <MAT>
 A;Cross-references: GB:D82884; NID:G2804261; PIDN:BAA24442.1; PID:G2804262
 C;Comment: This enzyme acts as a secreted or lysosomal proteinase. It has various funct
 y organs, and cleaning in the malpighian tubule.
 C;Superfamily: papain
 C;Keywords: cysteine proteinase; hydrolase
 F:145,284,305/Active site: Cys, His, Asn #status Predicted

Query Match 32.9%; Score 613.5; DB 2; Length 338;
 Best Local Similarity 43.6%; Pred. No. 1.2e-43;
 Matches 137; Conservative 49; Mismatches 107; Indels 21; Gaps 10;

Qy 56 FARFARRYKSYGSEEEIKKRFIGFVENLAFIRSTNR-----KDLSTYTLGINQFADLTWEE 111

Db 27 WSPFMQHSKNYDSFTEERFRMKIPWENAHKAKHKLPSQGFVKFGLNLYADMLHHE 86

Qy 112 FRTNLGAAQNCASATAGNH-----RFVDGV---LPVTRDWRQGIQVSPVKDQSCGS-W 162

Db 87 FVSTLNGFNKTNILKGSDDLNDVAFISPAVVKLPDVTVDWRDKGAVTEKVDQGHCGSCW 146

Qy 163 TFSSTTGALAAAYTQLTGS--TLSEQQLVDCASAFNFGC-GGLPSQAFVVKYNGGIDTE 219

Db 147 SFSATGSLGQHFRTKTKLVLSLEQNLVDCSGRYGNNCGNCGGLMDNNAFYIKDNGIDTE 206

Qy 220 QTYPYLGVVNGICNFQKQNVGVKVIDSINITLGADELKHAVALVPRVSVAFV 278

Db 207 KSVPYLADEKCHYKAQNSGATDKGFVDIEANEDDLKAAVATVGPVSIADASHETFOL 266

Qy 279 YKGVYSSDTQGRDPMVNHAVLAGYGV-EDGIPYWLKNSWGTNMGDNGYFKWELGK- 336

Db 267 YSDGVYSDEPCSSQLD--HGVLVVGYGTSDDQDYVLVKNWSGWSGLNGYIKWARRQD 324

Qy 337 NMCGVATCASYPIV 350

T12039
 cysteine proteinase (EC 3.4.22.-) 1 precursor - kidney bean
 C:Species: Phaseolus vulgaris (kidney bean)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
 C:Accession: T12039
 R:Senyuk, V.; Becker, C.; Muentz, K.
 submitted to the EMBL Data Library, October 1997
 A:Description: Isolation of cDNA clone encoding cysteine proteinase (Cp1) from a cotype

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 29, 2004, 21:10:38 ; Search time 25 Seconds
(without alignments)
733.148 Million cell updates/sec

Title: US-10-087-714-2
Perfect score: 1865
Sequence: 1 MAKKLFFLLVLSVAL.....ELGKMGVATCASYPIVAV 352

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1322.5	70.9	360	1	Q10717 zea mays (m
2	1281.5	68.7	362	1	P25778 oryza sativ
3	1274.5	68.3	362	1	P05167 hordeum vul
4	1249	67.0	356	1	Q40143 lycopersico
5	796	42.7	333	1	P00786 rattus norv
6	792	42.5	335	1	O46427 sus scrofa
7	765	41.0	333	1	P49935 mus musculu
8	759	40.7	335	1	P09668 homo sapien
9	609	32.7	462	1	P43297 arabidopsis
10	603.5	32.4	339	1	Q26636 sarcophaga
11	594.5	31.9	323	1	P25782 homarus ame
12	589	31.6	344	1	P54640 dictyosteli
13	587	31.5	471	1	P25777 oryza sativ
14	586	31.4	322	1	P13277 homarus ame
15	581	31.2	341	1	Q95029 drosophila
16	570	30.6	328	1	P25251 brassica na
17	563.5	30.2	363	1	P25804 pisum sativ
18	558	29.9	376	1	P04989 dictyosteli
19	556.5	29.8	321	1	P25784 homarus ame
20	556	29.8	380	1	P00785 actinidia c
21	554.5	29.7	458	1	P25776 oryza sativ
22	550	29.5	364	1	Q9aut0 arabidopsis
23	543.5	29.1	362	1	Q91t77 arabidopsis
24	541.5	29.0	334	1	O60911 homo sapien
25	539.5	28.9	371	1	Q9seu9 arabidopsis
26	539	28.9	361	1	P43295 arabidopsis
27	536.5	28.8	368	1	P43296 arabidopsis
28	531.5	28.5	329	1	O77641 macaca fasc
29	528.5	28.3	329	1	P43236 oryctolagus
30	528.5	28.3	333	1	Q91014 mus musculu
31	527.5	28.3	329	1	P43235 homo sapien
32	527.5	28.3	371	1	Q10716 zea mays (m
33	527	28.3	362	1	P12412 vigna mungo

ALIGNMENTS

RESULT 1

ID	CYS2 MAIZE	STANDARD;	PRT;	360 AA.
AC	Q10717;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Cysteine proteinase 2 precursor (EC 3.4.22.-).			
GN	CCP2.			
OS	Zea mays (Maize).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;			
OC	PACCAD clade; Panicoideae; Andropogoneae; Zea.			
OX	NCBI_TaxID=4577;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Seed;			
RX	MEDLINE=96004895; PubMed=7548211;			
RA	Domoto C., Watanabe H., Abe M., Abe K., Arai S.;			
RT	"Isolation and characterization of two distinct cDNA clones encoding			
RT	corn seed cysteine proteinases.";			
RL	Biochim. Biophys. Acta 1263:241-244(1995).			
CC	-1- FUNCTION: Involved in the degradation of the storage protein zein.			
CC	May play a role in proteolysis during emergencies.			
CC	-1- SUBCELLULAR LOCATION: Vacuolar (Probable).			
CC	-1- TISSUE SPECIFICITY: Expressed at the onset of germination.			
CC	-1- SIMILARITY: TO OTHER EUKARYOTIC THIOL PROTEASES.			
CC	-1- SIMILARITY: Belongs to peptidase family C1.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL; D45403; BAA08245.1; -			
DR	PIR; S59598; S59598.			
DR	HSNP; O46427; 8PCH.			
DR	MEDDB; 25431; -			
DR	MaizeDB; 25431; -			
DR	InterPro; IPR000169; SHprot acsite.			
DR	InterPro; IPR000169; SHprot acsite.			
DR	Pfam; PF00112; Peptidase_C1; 1.			
DR	PRINTS; PR00705; PAPAIN.			
DR	ProDom; PD000158; Peptidase_C1; 1.			
DR	SMART; SM00845; Pept_C1; 1.			
DR	PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.			
DR	PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.			
DR	PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.			
KW	Hydrolase; Thiol protease; Zymogen; Glycoprotein; signal.			
FT	SIGNAL 1 19 POTENTIAL.			
FT	PROPEP 20 142 ACTIVATION PEPTIDE (POTENTIAL).			
FT	CHAIN 143 360 CYSTEINE PROTEINASE 2.			
FT	ACT_SITE 167 167 BY SIMILARITY.			

O35186 rattus norv
P25803 phaseolus v
P15242 rattus norv
P25249 hordeum vul
Q9gie3 sus scrofa
Q28944 sus scrofa
P54639 dictyosteli
P25250 hordeum vul
O70370 mus musculu
P55097 mus musculu
P25975 bos taurus
Q9ubx1 homo sapien

FT ACT_SITE 307 307 BY SIMILARITY.
 FT ACT_SITE 327 327 BY SIMILARITY.
 FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 256 256 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 360 AA; 39199 MW; 8B5788F1B0C3FC1C CRC64;
 Query Match 70.9%; Score 1322.5; DB 1; Length 360;
 Best Local Similarity 71.7%; Pred. No. 7.4e-98;
 Matches 258; Conservative 28; Mismatches 65; Indels 9; Gaps 6;
 QY 1 MAAKLFFLLFVLSALVAL--AGFEEDNPIRSVTPRPS-IEPAILGLVGSCHAF 57
 DB 1 MVPRRLFLVAVVLDATAVAVNSGADSNPIRPTVDRASALESVFVAALGRTRDALRFA 60
 QY 58 RPARRYGKSGYSEEEIKKRGIFVENLAFTSTNRKDLSTYLGINQFADLTWEERTNRL 117
 DB 61 RFARYGKSGYSEAAVHKRIFSLSSQLVSTNRKGLSYRGLINRPFADMSWEEFRAFL 120
 QY 118 GAAQNCSATAGNHRF--VDGVLPTVRDREOGIVSPVKDQSCGS-WTFSTTGALAA 174
 DB 121 GAAQNCSATLTGNHRMRAAAVALPETKQWREDGIVSPVKNQCHGSCWTFSTTGALAA 180
 QY 175 TQLTGS--TLSEQQLVDCASAFNFGC--GGLPSQAFYKNGGIDTCTQTPYLGVMGIC 231
 DB 181 TQATGRPISEQLVDCGFAPNFGCNGGLPSQAFYKNGGIDTCTQTPYLGVMGIC 240
 QY 232 NFKQENVGVKVIDSINITLGADELKHAVALVPSVAPEVVKGNFKYKGYSSDTCGR 291
 DB 241 KFNENGVKVLDSNVITLGADELKDAVGLRVPVSAFEVITGRFLYKSGYTSDHCGT 300
 QY 292 DPMVDVNHAVLAVGYGVEDGIPWLIKNSWGTWNGDGYFKMELGKMGCVATCASYPI 351
 DB 301 TPDVNDVNHAVLAVGYGVEDGVPYWLKNSWGADWDGDEGYFKMGMKMGCVATCASYPI 360

RESULT 2

ORIC-ORISA STANDARD; PRT; 362 AA.
 ID ORIC-ORISA
 AC P25778;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Oryza gamma chain precursor (EC 3.4.22.-).
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare; TISSUE=Seed;
 RX MEDLINE=91358494; PubMed=1885617;
 RA Watanabe H., Abe K., Emori Y., Hosoyama H., Arai S.;
 RT "Molecular cloning and gibberellin-induced expression of multiple
 RL cysteine proteinases of rice seeds (oryzains).";
 RL J. Biol. Chem. 266:16897-16902(1991).
 CC -1- TISSUE SPECIFICITY: Expressed only in seeds.
 CC -1- INDUCTION: By gibberellic acid (GA).
 CC -1- SIMILARITY: Belongs to peptidase family C1.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D90408; BAA14404.1; -.
 DR PIR; J00390; KHRZOG.
 DR HSP; P07711; ICLJL.
 DR MEROPS; C01.041; -.
 DR Gramene; P25778; -.

DR InterPro; IPR000669; Peptidase_C1.
 DR InterPro; IPR000169; SHprot_accsite.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPAIA.
 DR ProDom; PD000158; Peptidase_C1; 1.
 DR SMART; SM00645; Pept_C1; 1.
 DR PROSITE; PS001139; THIOI_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOI_PROTEASE_HIS; 1.
 DR PROSITE; PS00640; THIOI_PROTEASE_ASN; 1.
 KW Hydrolase; Thiol protease; Zymogen; Glycoprotein; Signal.
 FT SIGNAL 24
 FT PROPEP 1 144 ACTIVATION PEPTIDE (POTENTIAL).
 FT CHAIN 145 362 ORYZAIN GAMMA CHAIN.
 FT ACT_SITE 169 169 BY SIMILARITY.
 FT ACT_SITE 309 309 BY SIMILARITY.
 FT ACT_SITE 329 329 BY SIMILARITY.
 FT DISULFID 166 209 BY SIMILARITY.
 FT DISULFID 200 242 BY SIMILARITY.
 FT DISULFID 300 350 BY SIMILARITY.
 FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 362 AA; 39204 MW; 73E573EE1DA10B9D CRC64;
 Query Match 68.7%; Score 1281.5; DB 1; Length 362;
 Best Local Similarity 66.9%; Pred. No. 1.3e-94;
 Matches 242; Conservative 45; Mismatches 64; Indels 11; Gaps 6;
 QY 1 MAAKLFFLLF-----VSALSVALAGFEEDNPIRSVTPRPS-IEPAILGLVGSCHAF 54
 DB 1 MAHRRITLLLAVAVAATSAVAASGDFDSDNPIRSVTDHAASALESTVIAALGRTRGAL 60
 QY 55 HFARFARYGKSGYSEEEIKKRGIFVENLAFTSTNRKDLSTYLGINQFADLTWEERT 114
 DB 61 RFARFARYGKSGYSEAAVHKRIFSLSSQLVSTNRKGLSYRGLINRPFADMSWEEFQA 120
 QY 115 NRLGAAQNCSATAGNHRFVDG-VLPVTRDREOGIVSPVKDQSCGS-WTFSTTGALAA 172
 DB 121 SRLGAAQNCSATLTAGNHRMEDAPALPETKQWREDGIVSPVKDQCHGSCWFFSTTGLEA 180
 QY 173 AYTQLTGS--TLSEQQLVDCASAFNFGC--GGLPSQAFYKNGGIDTCTQTPYLGVMG 229
 DB 181 RYTOATGPVPSLSBQQLADCATRYNFGCGGLPSQAFYKNGGIDTCTQTPYLGVMG 240
 QY 230 ICNFKQENVGVKVIDSINITLGADELKHAVALVPSVAPEVVKGNFKYKGYSSDTC 289
 DB 241 ICHYKPNAGVKVLDSNVITLGADELKHAVALVPSVAPEVVKGNFKYKGYSSDTC 300
 QY 290 GRDPMVDVNHAVLAVGYGVEDGIPWLIKNSWGTWNGDGYFKMELGKMGCVATCASYPI 349
 DB 301 GTSMDVNHAVLAVGYGVENGVPYWLKNSWGADWDGDEGYFKMGMKMGCVATCASYPI 360
 QY 350 VA 351
 DB 361 VA 362
 RESULT 3
 ALEU_HORVU
 ID ALEU_HORVU STANDARD; PRT; 362 AA.
 AC P05167;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Thiol protease aleurain precursor (EC 3.4.22.16).
 OS Hordeum vulgare (Barley).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
 OC Triticeae; Hordeum.
 OX NCBI_TaxID=4513;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86016732; PubMed=3901004;
 DR Rogers J.C., Dean D., Heck G.R.;

"Aleurain: a barley thiol protease closely related to mammalian cathepsin H.",
Proc. Natl. Acad. Sci. U.S.A. 82:6512-6516 (1985).
(2)

RP REVISIONS.

RA Rogers J.C.;
Submitted (MAR-1987) to the EMBL/GenBank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: Hydrolysis of proteins, acting as an
aminopeptidase (notably, cleaving Arg-|-Xaa bonds) as well as an
endopeptidase.

CC -1- SUBCELLULAR LOCATION: VACUOLE-LIKE SUBCELLULAR COMPARTMENT.

CC -1- INDUCTION: Aleurain is synthesized by the aleurone cells

CC stimulated by gibberellic or abscisic acid.

CC -1- SIMILARITY: Belongs to peptidase family C1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; X05167; CAA28804.1; -.

DR HSSP; O46427; 8PCH.

DR MEROPS; C01.041; -.

DR InterPro; IPR000668; Peptidase C1.

DR InterPro; IPR000169; SHprot.acSite.

DR Pfam; PF00112; Peptidase_C1; 1.

DR PRINTS; PR00705; PAPAIN.

DR ProDom; PD000158; Peptidase_C1; 1.

DR SMART; SM00645; Pept C1; 1.

DR PROSITE; PS00139; THIOI PROTEASE CYS; 1.

DR PROSITE; PS00639; THIOI PROTEASE_HIS; 1.

DR PROSITE; PS00640; THIOI PROTEASE_ASN; 1.

KW Hydrolase; Thiol protease; Glycoprotein; Zymogen; Signal; Germination.

FT SIGNAL 1 22

FT PROPEP 23 143 ACTIVATION PEPTIDE (POTENTIAL).

FT CHAIN 144 362 THIOI PROTEASE ALEURAIN.

FT ACT_SITE 168 168 BY SIMILARITY.

FT ACT_SITE 308 308 BY SIMILARITY.

FT ACT_SITE 328 328 BY SIMILARITY.

FT DISULFID 165 208 BY SIMILARITY.

FT DISULFID 199 241 BY SIMILARITY.

FT DISULFID 299 349 BY SIMILARITY.

FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SIGNAL 257 257

FT SEQUENCE 362 AA; 39122 MW; A70CCD4A843A1686 CRC64;

Query Match 68.3%; Score 1274.5; DB 1; Length 362;

Best Local Similarity 67.9%; Pred. No. 4.9e-94;

Matches 243; Conservative 42; Mismatches 64; Indels 9; Gaps 6;

3 AKLLFFLLFLVLSVALA---GPEENPIRSVTQRPDS-IEPAILGVLSGCRHAFHARF 58

4 ARVLLALAVLATAAVALASSSSFADSNIPVPTDRAASTLESVALGALGRTHALREFAR 63

59 PARYKSGYSGEIEIKRFGIFVFNLAFTSTNRKOLSYVLGINOPADLTWEEFRTNRLG 118

64 FAVRYKSGYSAEAVERRRFRIFSELEVRSTNRKGLPYRLGINRFSMDSWEEFOATRLG 123

119 AAOCSATAHGNHRFVD-GVLPVTRDREOGIVSPVKDQSCGS-WTFSTTGALEAAATQ 176

124 AAQTCSATLGNHLMRDAAALPEYKDNREDGIVSPVKNQAHGSCWTFSTTGALEAAATQ 183

177 LTGS--TLSEQQLVDCASAFNFGC-GGLPSQAFYVKYNGGIDTETQTYPLGVMGICNF 233

184 ATGNKISLSEQLVDCAGGFNFGCGLPSQAFYKNGGIDTETSYPKYGVNGVCHY 243

234 KQENVGVKVIDSINITGADELKHAVGLRVPVSAFEVVKGNLYKKGYYSDTCGRDP 293

244 KAENAAQVLDVSNITLNAEDELKNAVGLRVPVSAFQVLDGFRQYKSGVYTSDHCGTTP 303

QY 294 MDVNEAVLAVGYGVEDGIPYWLKNSWGTNWDNGYFKMELGKNMCGVATCASYPIVA 351

DB 304 DDVNEAVLAVGYGVENGYPYWLKNSWGDWNGYFKMELGKNMCAATATCASYPWA 361

RESULT 4

CYS3_LYCES

ID CYS3_LYCES STANDARD; PRT; 356 AA.

AC Q40143;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Cysteine proteinase 3 precursor (EC 3.4.22.-).

GN CYP-3.

OS Lycopersicon esculentum (Tomato).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;

OC Lamiales; Solanales; Solanaceae; Solanum.

OX NCBI_TaxID=4081;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Ailsa Craig; TISSUE=Leaf;

RX MEDLINE=96194458; PubMed=8624407;

RA Drake R., John I., Farrell A., Cooper W., Schuch W., Grierson D.;

RT "Isolation and analysis of cDNAs encoding tomato cysteine proteases

RT expressed during leaf senescence.";

RL Plant Mol. Biol. 30:755-767(1996).

CC -1- SUBCELLULAR LOCATION: Vacuolar (Probable).

CC -1- TISSUE SPECIFICITY: Predominantly expressed in stem and root.

CC -1- SIMILARITY: Belongs to peptidase family C1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; Z48736; CAA86629.1; -.

DR PIR; S66348; S66348.

DR HSSP; O46427; 8PCH.

DR MEROPS; C01.041; -.

DR InterPro; IPR000668; Peptidase C1.

DR InterPro; IPR000169; SHprot.acSite.

DR Pfam; PF00112; Peptidase_C1; 1.

DR PRINTS; PR00705; PAPAIN.

DR ProDom; PD000158; Peptidase_C1; 1.

DR SMART; SM00645; Pept C1; 1.

DR PROSITE; PS00139; THIOI PROTEASE CYS; 1.

DR PROSITE; PS00639; THIOI PROTEASE_HIS; 1.

DR PROSITE; PS00640; THIOI PROTEASE_ASN; 1.

KW Hydrolase; Thiol protease; Glycoprotein; Zymogen; Signal.

FT SIGNAL 1 16

FT PROPEP 17 138 ACTIVATION PEPTIDE (POTENTIAL).

FT CHAIN 139 356 CYSTEINE PROTEINASE 3.

FT ACT_SITE 163 163 BY SIMILARITY.

FT ACT_SITE 303 303 BY SIMILARITY.

FT ACT_SITE 323 323 BY SIMILARITY.

FT DISULFID 160 203 BY SIMILARITY.

FT DISULFID 194 236 BY SIMILARITY.

FT DISULFID 294 349 BY SIMILARITY.

FT CARBOHYD 123 123 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SEQUENCE 356 AA; 38944 MW; BE176A7C74BCDD44 CRC64;

Query Match 67.0%; Score 1249; DB 1; Length 356;

Best Local Similarity 68.3%; Pred. No. 5.1e-92;

Matches 243; Conservative 35; Mismatches 70; Indels 8; Gaps 5;

3 AKLLFFLLFLVLSVALAG---FEENPIRSVTQRPDSIEPAILGVLSGCRHAFHARF 59

2 SRLSLVLVLAVGLATAGPATFADKNPIROV-FPELENGILQVVGQTRSAISAF 60

RESULT 6

CATH_PIG STANDARD; PRT; 335 AA.

AC O46427; 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Cathepsin H precursor (EC 3.4.22.16).

GN CTSR.

OS Sus scrofa (Pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

OX NCBI_TaxID=9823;

RN [1]

RP SEQUENCE FROM N.A., AND X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).

RC TISSUE=Peripheral blood, and Spleen;

RX MEDLINE=98154318; PubMed=9493267;

RA Guncar G., Podobnik M., Pungercar J., Strukelj B., Turk V., Turk D.,

RT "Crystal structure of porcine cathepsin H determined at 2.1-A

RT resolution; location of the mini-chain C-terminal carboxyl group

RT defines cathepsin H aminopeptidase function.";

RL Structure 6:51-61(1998).

CC -!- FUNCTION: Important for the overall degradation of proteins in lysosomes.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins, acting as an aminopeptidase (notably, cleaving Arg-Xaa bonds) as well as an endopeptidase.

CC -!- SUBUNIT: Composed of a minichain and a large chain. The large chain may be split into heavy and light chain. All chains are held together by disulfide bonds.

CC -!- SUBCELLULAR LOCATION: Lysosomal.

CC -!- SIMILARITY: Belongs to peptidase family C1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).

CC -----

DR EMBL; AF001169; AAB93957.1; -.

DR PDB; 8PCF; 13-JAN-99.

DR PDB; 1NB3; 18-FEB-03.

DR MEROPS; C01.040; -.

DR InterPro; IPR000568; Peptidase C1.

DR InterPro; IPR000169; SHprot_acsite.

DR Pfam; PF00112; Peptidase_C1; 1.

DR PRINTS; PR00705; PAPAIN.

DR ProDom; PD000158; Peptidase_C1; 1.

DR SMART; SM00645; Pept C1; 1.

DR PROSITE; PS00139; THIOLEPROTEASE_CYS; 1.

DR PROSITE; PS00639; THIOLEPROTEASE_HIS; 1.

DR PROSITE; PS00840; THIOLEPROTEASE_ASN; 1.

KW Hydrolase; Thiol protease; Lysosome; Glycoprotein; Zymogen; Signal;

KW 3D-structure.

FT SIGNAL 1 22 POTENTIAL.

FT PROPEP 23 97 POTENTIAL.

FT CHAIN 98 105 CATHEPSIN H MINICHAIN.

FT PROPEP 107 115

FT CHAIN 116 335 CATHEPSIN H.

FT ACT_SITE 141 141

FT ACT_SITE 281 281

FT ACT_SITE 301 301

FT DISULFID 102 327

FT DISULFID 138 181

FT DISULFID 172 214

FT DISULFID 272 322

FT CARBOHYD 72 72

FT CARBOHYD 101 101

FT CARBOHYD 230 230

N-LINKED (GLCNAC. . .) (POTENTIAL).

N-LINKED (GLCNAC. . .) (POTENTIAL).

N-LINKED (GLCNAC. . .) (POTENTIAL).

FT STRAND 120 121

FT HELIX 122 125

FT STRAND 134 134

FT TURN 136 137

FT HELIX 141 158

FT STRAND 164 164

FT HELIX 166 172

FT TURN 177 177

FT HELIX 180 182

FT HELIX 186 196

FT TURN 197 197

FT STRAND 199 201

FT TURN 202 204

FT STRAND 218 220

FT STRAND 221 223

FT TURN 225 230

FT HELIX 233 234

FT TURN 236 245

FT STRAND 246 246

FT STRAND 249 253

FT HELIX 257 260

FT STRAND 266 267

FT TURN 276 278

FT STRAND 281 291

FT TURN 292 293

FT STRAND 294 300

FT TURN 303 306

FT TURN 308 308

FT STRAND 309 309

FT TURN 310 311

FT STRAND 312 315

FT HELIX 321 323

FT TURN 324 326

FT STRAND 329 333

SQ SEQUENCE 335 AA; 37454 MW; F728E845292C3B55 CRC64;

Query Match 42.5%; Score 792; DB 1; Length 335;

Best Local Similarity 54.3%; Pred. No. 9.9e-56;

Matches 165; Conservative 37; Mismatches 94; Indels 8; Gaps 7;

QY 55 HFARFARYKSYGSEEEIKKRGIFVEMLAIFRSTNRKDLSTLTGINOFADLTWEEFRT 114

DB 34 HFKSMVMVQKKY-SLEEYHRLQVPSVNRKINAHNAGNHTPKLGLNQFSDMSFDEIRH 92

QY 115 NRI-GAAQNCATAHGNHRFVQGLVPTDRDREQ-IVSPVKDQSGCS-WTFSTTTGALE 171

DB 93 KYLWSEFQNCAT-KGYLRGTGTPYPPSMDWRKKNFVSPVKNQSGCGSCWTFSTTTGALE 151

QY 172 AAYTQLTGSTLS--EQQLVDCASAFNNFC-GGLPSQAFYKYNKGIDITEQTPYLVGM 228

DB 152 SAVAIATGKMLSLAEQQLVDCAQNFNNHCGQGLPSQAFYIRYNKGIMGEDTYPYKQD 211

QY 229 GICNFQENGVGVKVIDSINITLGADELKHAQLVRPSVAPEVVKGNLYKKGVYSSDT 288

DB 212 DHCKFPDKAIAFVKDVAITWDEAMVEALVNPVSAFEVTNDFLMIRKGIYSSTIS 271

QY 289 CGRDPMDVNHAVLAVGYGVEDGIPYWLIIKNSMGTWMDNGYFPMELGKNMCGVATCASP 348

DB 272 CHKTPDKVNHAVLAVGYGEENGIPYWIKNKSWGQWGMNGYFLIERGKNMCGLAACASP 331

QY 349 IVAV 352

DB 332 IPLV 335

RESULT 7

CATH_MOUSE STANDARD; PRT; 333 AA.

ID CATH_MOUSE

AC P49935;

DT 01-OCT-1996 (Rel. 34, Created)

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 88-335 FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=89076480; PubMed=2849458;
RA Fuchs R., Machleidt W., Gassen H.G.;
RT "Molecular cloning and sequencing of a cDNA coding for mature human
RT kidney cathepsin H";
RL Biol. Chem. Hoppe-Seyler 369:469-475(1988).
RN [4]
RP SEQUENCE OF 98-105 AND 114-335.
RC TISSUE=Kidney;
RX MEDLINE=88137635; PubMed=2342889;
RA Ritonja A., Popovic T., Kotnik M., Machleidt W., Turk V.;
RT "Amino acid sequences of the human kidney cathepsins H and L";
RL FEBS Lett. 228:341-345(1988).
RN [5]
RP SEQUENCE OF 99-105; 116-159 AND 294-335.
RA Machleidt W., Ritonja A., Popovic T., Kotnik M., Brzin J., Turk V.,
RA Machleidt I., Muller-Esterl W.;
RT "Human cathepsins B, H and L: Characterization by amino acid sequences
RT and some kinetics of inhibition by the kininogens";
RL (In) Turk V. (eds.);
RL Cysteine proteinases and their inhibitors, pp.13-18,
RL Walter de Gruyter, Berlin and New York (1986).
CC -1- FUNCTION: Important for the overall degradation of proteins in
CC lysosomes.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of proteins, acting as an
CC aminopeptidase (notably, cleaving Arg-|-Xaa bonds) as well as an
CC endopeptidase.
CC -1- SUBUNIT: Composed of a minichain and a large chain. The large
CC chain may be split into heavy and light chain. All chains are
CC held together by disulfide bonds.
CC -1- SUBCELLULAR LOCATION: Lysosomal.
CC -1- SIMILARITY: Belongs to peptidase family C1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X07549; CAA30428.1; -;
CC EMBL; X07549; CAA30429.1; ALT_SEQ.
CC EMBL; BC002479; AA02479.1; -;
CC EMBL; X16832; CAA34734.1; -;
CC PIR; S12486; KKHUH.
CC PRB; 1BZN; 18-OCT-99.
CC MEROPS; C01.040; -;
CC Genew; HGNC:2535; CTSH.
CC MIM; 116820; -;
CC GO; GO:0004215; F:cathepsin H activity; TAS.
CC GO; GO:0008234; F:cysteine-type peptidase activity; TAS.
CC GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
CC InterPro; IPR000668; Peptidase_C1.
CC InterPro; IPR000169; Shprot_acsite.
CC Pfam; PF00112; Peptidase_C1; 1.
CC PRINTS; PR00705; PAPAIN.
CC ProDom; PD000158; Peptidase_C1; 1.
CC SMART; SM00645; Pept_C1; 1.
CC PROSITE; PS00139; THIOI_PROTEASE_CYS; 1.
CC PROSITE; PS00639; THIOI_PROTEASE_HIS; 1.
CC PROSITE; PS00640; THIOI_PROTEASE_ASN; 1.
CC Hydrolase; Thiol protease; Lysosome; Glycoprotein; Zymogen; Signal;
CC 3D-structure.
CC SIGNAL
CC PROPEP 1 22 97 ACTIVATION PEPTIDE.
CC FT

FT PEPTIDE 98 105 CATHEPSIN H MINICHAIN.
FT CHAIN 116 335 CATHEPSIN H.
FT CHAIN 116 292 CATHEPSIN H HEAVY CHAIN.
FT CHAIN 293 335 CATHEPSIN H LIGHT CHAIN.
FT ACT_SITE 141 141 BY SIMILARITY.
FT ACT_SITE 281 281 BY SIMILARITY.
FT ACT_SITE 301 301 BY SIMILARITY.
FT DISULFID 102 327 BY SIMILARITY.
FT DISULFID 138 181 BY SIMILARITY.
FT DISULFID 172 214 BY SIMILARITY.
FT DISULFID 272 322 BY SIMILARITY.
FT CARBOHYD 230 230 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT CONFLICT 179 179 H -> Y (IN REF. 1 AND 2).
FT CONFLICT 306 306 Q -> E (IN REF. 4 AND 5).
SQ SEQUENCE 335 AA; 37377 MW; 4690615EE155B767 CRC64;
Query Match 40.7%; Score 759; DB 1; Length 335;
Best Local Similarity 52.1%; Pred. No. 4.2e-53;
Matches 159; Conservative 39; Mismatches 99; Indels 8; Gaps 7;
QY 54 FHFARFARYKSYGSBEEIKRFGIFVENIAFIRSTNRKDLSTLTGINQADLTWEFR 113
DB 33 FHPKSWMSKHKTY-STEEVHRLQTFASNRKINAHNGNHTFKMALNQSDMSFABIK 91
QY 114 TNRI-GRAONCSATAGNHRFVDGVLPTDRDREQ-IVSPVKDQSGS-WTFSTTGAL 170
DB 92 HKYLWSEPNCSAT-KSNILRGTPYPSVDNRKGNFVSPVKQAGCGSWFTTGAL 150
QY 171 EAAVTQLTGTSLIS--EQLVDCASAFNFGC-GGLPSQAFYVYKNGSIDTEQTPYLG 227
DB 151 ESAIATATGKMLSLAEQQLVDCADFNHGGCGGLPSQAFYILYKNGIMGEDTPYQ 210
QY 228 MGICNFQKQVGVKVIDSINITLGADELKHVGLVPSVAVFVVKGFNLYKGVYSSD 287
DB 211 DGYCKFGPKGAIGFVKDVANITIIYDEAWVEAVALYNPVSFAFVTDQFMRYRTGYSST 270
QY 288 TCGRDPDMVNHAVLAVGYGVEDGIPYMLIKNSWGTNGDNGYFKNELGKNMCGVATCASY 347
DB 271 SCHKTPDKVNHAVLAVGYGKNGIPYMLVKNWGPQGMNGYFLIERKNNMCGLAACASY 330
QY 348 PIVAV 352
DB 331 PIPLV 335
RESULT 9
RD21 ARATH STANDARD; PRT; 462 AA.
ID RD21 ARATH
AC P43297;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cysteine proteinase RD21a precursor (EC 3.4.22.-) (RD21).
GN RD21A OR AT1647128 OR P2G19.31.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=93314960; PubMed=8325504;
RA Koizumi M., Yamaguchi-Shinozaki K., Tsuji H., Shinozaki K.;
RT "Structure and expression of two genes that encode distinct drought-
RT inducible cysteine proteinases in Arabidopsis thaliana";
RL Gene 129:175-182(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,

RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
 RA Mollisier J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Uterback T., Van Aken S., Vayberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.,
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 408:816-820(2000).
 CC -1- INDUCTION: By high salt conditions.
 CC -1- SIMILARITY: Belongs to peptidase family C1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D13043; BAA02374.1; -;
 DR EMBL: AC083835; AAG50628.1; -;
 DR F1R: JN0719; JN0719.
 DR HSSP: P00785; 2ACT.
 DR MEROPS: C01.029; -;
 DR InterPro: IPR000118; Granulin.
 DR InterPro: IPR000568; Peptidase_C1.
 DR InterPro: IPR000189; SHP100_acid.
 DR Pfam: PF00396; granulin; 1.
 DR PRINTS: P00705; PAPAIN.
 DR ProDom: PD000158; Peptidase_C1; 1.
 DR SMART: SM00277; GRAN; 1.
 DR SMART: SM00645; Rept C1; 1.
 DR PROSITE: PS00139; THIOL PROTEASE CYS; 1.
 DR PROSITE: PS00639; THIOL PROTEASE HIS; 1.
 DR PROSITE: PS00640; THIOL PROTEASE ASN; 1.
 KW Hydroxylase; Thiol protease; Zymogen; Glycoprotein; Signal.
 FT SIGNAL 1 21
 FT PROPEP 22 136
 FT CHAIN 137 462
 FT ACT_SITE 161 161
 FT ACT_SITE 297 297
 FT ACT_SITE 317 317
 FT DISULFID 158 200
 FT DISULFID 192 233
 FT DISULFID 291 342
 FT CARBOHYD 90 90
 FT CARBOHYD 414 414
 SQ SEQUENCE 462 AA; 50966 MW; 4C59955CCB95A58 CRC64;
 Query Match 32.7%; Score 609; DB 1; Length 462;
 Best Local Similarity 38.6%; Pred No. 5.1e-41;
 Matches 140; Conservative 73; Mismatches 112; Indels 38; Gaps 16;
 QY 5 LIFLLFLV-SALSVALAGFEDNPIRSVTPQPSIEPAILGVLSGRHAFHFPARRY 63
 Db 10 ILFLAMVSSAVDMSIISDEKKGVSSTGGRS---EAEVMSI-----YEALVVKH 57
 QY 64 GKSYSGEIEIKK--RFGIFVENLAFIRSTNRKLSYTLGINOPADLTWEFRNRLGAAQ 121
 Db 58 GKAQSONSLVEKDRFEKFKNLRVDENKNSYRLGTRADLTNDYRSKYLCAQM 117
 QY 122 NCSATAGNHRRF---VDGVLPVTRDREQGIVSPVKQGGCGS-WPSTTTGALEAYTOL 177

Db 118 EKGERTSRYEARVDELPESIDWRKKGVAEVRKQGGCGSWAFSTIGAVEGINIV 177
 QY 178 TGS--TLSEQLVDCASAFNFC--GGLPSQAFYKYNKGIDTBTQTPYLGWMGICNFK 234
 Db 178 TGLDLITLSEQLVDCQTSYNE--CGNGLMDYAFETIKNGGIDTDKDYPKGVGDTCDQI 236
 QY 235 QENVGVKVIDSI--NITLGAEDLKHAVGLVRVSVAFEV--VKGENLYKKGYSSTTCGRD 292
 Db 237 RKNKVVITSDYEDVPTYSSESLKAVA--HQPISTAIETAGRAFQLYDSGIFDG--SCG-- 292
 QY 293 PMDVNHAVLAVGVGVEDGIPYMLIKNSWTNWDNGYFPMELGKM-----CGVATCAS 346
 Db 293 -TQLDHGWAAGVGTENGKDYMTVRNSWKGSGESGLRM--ARNIASSSGKCGIAIBPS 349
 QY 347 YPI 349
 Db 350 YPI 352
 RESULT 10
 CATL_SARPE
 ID CATL_SARPE STANDARD; PRT; 339 AA.
 AC Q26636;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cathepsin L precursor (EC 3.4.22.15).
 OS Sarcophaga peregrina (flesh fly) (Boettcherisca peregrina).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Oestroidea;
 OC Sarcophagidae; Sarcophaga.
 NCBI_TaxID:7386;
 RN NCBI
 RX NCB
 RP SEQUENCE FROM N.A., SEQUENCE OF 18-27; 122-135; 139-154; 200-208;
 RP 224-239 AND 306-319, FUNCTION, SUBCELLULAR LOCATION, AND DEVELOPMENTAL
 RP STAGE;
 RX MEDLINE=94253090; PubMed=8195162;
 RA Homma K.-I., Kurata S., Natori S.;
 RT "Purification, characterization, and cDNA cloning of procathepsin L
 RT from the culture medium of NIH-Sape-4, an embryonic cell line of
 RT sarcophaga peregrina (flesh fly), and its involvement in the
 RT differentiation of imaginal discs";
 RL J. Biol. Chem. 269:15258-15264(1994).
 CC -1- FUNCTION: Important for the overall degradation of proteins in
 CC lysosomes. Required for differentiation of imaginal
 CC disks.
 CC -1- CATALYTIC ACTIVITY: Specificity close to that of papain. As
 CC compared to cathepsin B, cathepsin L exhibits higher activity
 CC towards protein substrates, but has little activity on L-Arg-Arg-
 CC -NMEC- and no peptidyl-dipeptidase activity.
 CC -1- SUBUNIT: Dimer of a heavy and a light chain linked by disulfide
 CC bonds (by similarity).
 CC -1- SUBCELLULAR LOCATION: Lysosomal.
 CC -1- DEVELOPMENTAL STAGE: Highly expressed during embryonic development
 CC with higher levels in first instar than in third instar.
 CC -1- SIMILARITY: Belongs to peptidase family C1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D16533; BAA03970.1; -;
 DR PIR: A53810; A53810.
 DR HSSP: P07711; 1CJL.
 DR MEROPS: C01.092; -;
 DR InterPro: IPR000668; Peptidase_C1.
 DR InterPro: IPR000169; SHP100_acid.
 DR Pfam: PF00112; Peptidase_C1; 1.
 DR PRINTS: P00705; PAPAIN.

```

DR ProDom; PD000158; Peptidase_C1; 1.
DR SMART; SM00645; Pept_C1; 1.
DR PROSITE; PS00139; THIOLESTERASE_CYS; 1.
DR PROSITE; PS00639; THIOLESTERASE_HIS; 1.
DR PROSITE; PS00640; THIOLESTERASE_ASN; 1.
KW Hydrolase; Thiol protease; Glycoprotein; Lysosome; Zymogen; Signal;
KW Differentiation; Developmental protein.
FT SIGNAL 1 17
FT PROPEP 18 121
FT CHAIN 122 294
FT PROPEP 295 339
FT CHAIN 299 339
FT ACT_SITE 146 146
FT ACT_SITE 285 285
FT ACT_SITE 306 306
FT DISULFID 143 186
FT DISULFID 177 219
FT DISULFID 278 328
FT CARBOHYD 96 96
SQ SEQUENCE 339 AA; 37847 MW; 7401F3281A2FDA33 CRC64;

Query Match 32.4%; Score 603.5; DB 1; Length 339;
Best Local Similarity 42.3%; Pred. No. 9.7e-41;
Matches 134; Conservative 48; Mismatches 96; Indels 39; Gaps 11;

QY 62 RYKGYGSEEEIKRFGIFVENLAFIRSTN---RKLSYTLGINOPADLTWEEF----- 112
DB 34 QKRYANEVEFRFMKIFENNRHIAKHNLFAQGVSYKGLNKKYADMLHHEFKETMN 93
QY 113 -----RTNLGAQNCATAGHNRHFDVGLVPTVRDREQGVSPVKDQSGC 160
DB 94 GYNHTLRQMRERTGLVATY--IPPAH-----VTVPKSVDRHGAVTGKVDQGHG 144
QY 161 S-WTSTTCALAAVYTLTG--STLSEQLVDCASAFNNFGC-GGLPSQAFYVYKNGGI 215
DB 145 SCWAFSTSGALEGQFRKAGVLVLSQNLVDCSTKYGNNGGGLMDNATRYIKDNGGI 204
QY 217 DTEQYYPYLVGMGICNFQKQENVGVKVIDSINITLGADELKHAVGLRVPVSAFEVW-KG 275
DB 205 DTEKSYPEGIDDSCHFNFKATIGATDTGFDVPEGDEEKKKAVATMGFPVSAIDASHES 264
QY 276 FNLYKKGYSSTDCGRDPMVNHAVALVGYGV-EDGIPYWLKNSWNTWNGDNGYFMEL 334
DB 265 POLYSEGVNPEEC--DSQNLHDHGLVYGVGTDSGMDYLVKNSWNTWNGDNGYFMEL 322
QY 335 GK-NMCGVATCASYPIV 350
DB 323 NNQNCGIATASSYPTV 339

RESULT 11
ID_CYS2_HOMAM STANDARD; PRT; 323 AA.
AC P25782;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Digestive cysteine proteinase 2 precursor (EC 3.4.22.-).
GN LCP2.
OS Homarus americanus (American lobster).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Nephropoidea; Nephropidae; Homarus.
OX NCBI_TaxID=6706;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Digestive gland;
RX MEDLINE=92070467; PubMed=1959590;
RA Laycock M.V., Mackay R.M., di Fruescio M., Gallant J.W.;
RT "Molecular cloning of three cDNAs that encode cysteine proteinases in
the digestive gland of the American lobster (Homarus americanus).";
RL FEBS Lett. 292:115-120(1991).
CC -!- ENZYME REGULATION: Inhibited by E-64, antipain, leupeptin, heavy

```

```

CC metal ions, iodoacetic acid, dithionitrobenzene, p-hydroxymercuri-
CC benzoate; activated by mercaptoethanol and dithiothreitol.
CC -!- SIMILARITY: Belongs to peptidase family C1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X63568; CAA45128.1; -
DR PIR; S19650; S19650.
DR HSP; O60911; IPH0.
DR MEROPS; C01.030; -
DR InterPro; IPR000668; Peptidase_C1.
DR InterPro; IPR000169; SHprot acsite.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPAIN.
DR ProDom; PD000158; Peptidase_C1; 1.
DR SMART; SM00645; Pept_C1; 1.
DR PROSITE; PS00139; THIOLESTERASE_CYS; 1.
DR PROSITE; PS00639; THIOLESTERASE_HIS; 1.
DR PROSITE; PS00640; THIOLESTERASE_ASN; 1.
KW Hydrolase; Thiol protease; Zymogen; Signal; Multigene family.
FT SIGNAL 1 15
FT PROPEP 17 107
FT CHAIN 108 323
FT ACT_SITE 131 131
FT ACT_SITE 270 270
FT ACT_SITE 290 290
FT DISULFID 128 171
FT DISULFID 162 204
FT DISULFID 263 312
SQ SEQUENCE 323 AA; 35401 MW; 1AD93FC73C6E9831 CRC64;

Query Match 31.9%; Score 594.5; DB 1; Length 323;
Best Local Similarity 43.2%; Pred. No. 4.8e-40;
Matches 131; Conservative 40; Mismatches 119; Indels 13; Gaps 8;

QY 59 FARRYKSGSBEIEIKRFGIFVENLAFIRSTNTRK-----DLSYTLGINOPADLTWEEFRT 114
DB 23 FKGYGRQVDAEDSYRRVIFEQNKQYIEEFNKYENGVEVFNLANRFGDMTLEEFNA 82
QY 115 NRLGAQNCATAGHNRH-RFVDGLVPTVRDREQGVSPVKDQSGC-WTSTTCALAA 172
DB 83 VMKGNTPRSAPVSVFYPKKTGPQATEVDWRTKGAVTVPKDGQCGSCWAFSTTGLSG 142
QY 173 AVTQLTGS--TLSEQLVDCASAFNNFGC-GGLPSQAFYVYKNGGIDTEQYYPYLVGMG 229
DB 143 QHFLKGTGSLSLAEQQLVDCSRPYGPGCGNGGMMNDADFYIKANNNGIDTEAAYPYEAR 202
QY 230 ICNFKQENVGVKVIDSINITLGADELKHAVGLRVPVSAFEVW-KGFNLYKKGYSST 288
DB 203 SCRFDSNSVAATCSGHTNIASSETGLQAVRDIGIPSTIDAAKSSFOYSGVYFES 262
QY 289 CGRDPMDVNHAVALVGYGVEDGIPYWLKNSWNTWNGDNGYFMELGK-NMCGVATCASY 347
DB 263 C--SPSYLDHVALVGYGSEGGQDFWLKNSWNTWNGDNGYFMELKNSRNRNNGCIATVASY 320
QY 348 PIV 350
DB 321 PLV 323

RESULT 12
ID_CYS5_DICDI STANDARD; PRT; 344 AA.
AC P54640;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)

```

Cysteine proteinase 5 precursor (EC 3.4.22.-).

CPRE OR CPS.

Ddictyostelium discoideum (slime mold).

OCC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.

OCC NCBI_TaxID=44689;

(1) _____ RN

SEQUENCE FROM N.A.

STRAIN=AX4;

MEDLINE=96081966; PubMed=7499424;

Souza G.M., Hirai J., Mehta D.P., Freeze H.H.;

"Identification of two novel Dictyostelium discoideum cysteine

proteinases that carry N-acetylglucosamine-1-P-modification.";

J. Biol. Chem. 270:28938-28945 (1995).

-!- SUBCELLULAR LOCATION: Lysosomal.

-!- DEVELOPMENTAL STAGE: Present in the vegetative phase and decreases

with the start development, reappears in low levels when the

fruiting body is formed.

-!- PTM: Phosphoglycosylated, contains GlcNAc-alpha-1-P-Ser residues.

-!- SIMILARITY: Belongs to peptidase family C1.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL Outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL; L36205; AAA92018.1; --

HSSP; P07711; 1CUL.

MEROPS; C01.081; --

DictyBase; DB00168251; cprE.

InterPro; IPR000668; Peptidase_C1.

InterPro; IPR000169; SHprot_acsite.

Pfam; PF00112; Peptidase_C1; 1.

PRINTS; PR00705; PAPAIN.

ProDom; PD000158; Peptidase_C1; 1.

SMART; SMO0645; Pept C1; 1.

PROSITE; PS00139; THIOI_PROTEASE_CYS; 1.

PROSITE; PS00639; THIOI_PROTEASE_HIS; 1.

PROSITE; PS00640; THIOI_PROTEASE_ASN; 1.

Hydrolase; Thiol protease; Lysozyme; Zymogen; Glycoprotein;

phosphorylation; Signal.

SIGNAL

PROPEP 1 17 POTENTIAL.

CHAIN 18 111 ACTIVATION PEPTIDE (POTENTIAL).

DOMAIN 112 344 CYSTEINE PROTEINASE 5.

SER-RICH.

ACT_SITE 136 136 BY SIMILARITY.

ACT_SITE 272 272 BY SIMILARITY.

ACT_SITE 311 311 BY SIMILARITY.

DISULFID 133 174 BY SIMILARITY.

DISULFID 167 207 BY SIMILARITY.

DISULFID 265 333 BY SIMILARITY.

CARBOHYD 110 110 N-LINKED (GLCNAC...) (POTENTIAL).

CARBOHYD 297 297 N-LINKED (GLCNAC...) (POTENTIAL).

SEQUENCE 344 AA; 32712 MW; 82F3F5B8C147BA8 CRC64;

Query Match 31.6%; Score 589; DB 1; Length 344;

Best Local Similarity 36.7%; Pred. No. 1.4e-39;

Matches 139; Conservative 52; Mismatches 120; Indels 68; Gaps 13;

4 KLLFELLVLSVALAGREDNPIRSVTQRPDSIEPAILGVLCGRHAFHFARRARY 63

2 KVLSEFLVLSVATKQRSE -----LOYENAFDTMWTH 37

64 GKSGYSEBEIKRGIFVENLAFTSTRNRKDLVYLGINOFADLTWEEFRTNLGAQNC 123

38 QKSQYTS-EFGARGYNIFTANNDYVQQNSKGSETLVGLGNFAFDITNEEVNTLYGKFDA 96

124 SA-----TAGHNHRFVDGLPYLTRDWREQIVSPVKQCGSCGS-WTFSTTGAAEAAYT 175

97 SSLIGTQEKVHTNSS-----AASKDWRSEGAVTPVNQGCGGCSWFSTTGEGAHF 150

176 QLTS--TLSEQLVDCASAFNNFGC-GGLPSQAFFVFKNGCIDTEQYPPLGVWGICN 232

151 QSKGELVSLSEONLDCST--ENSQGGGLMTYAFYIIINNNDITDESSYPYKAENGKCE 208

233 FKQENVGVKVIDSINITLGAEDELKHAVGLRVFSVAFEVV-KGFNLKYKKGVSSDTCGR 291

209 YKSENSGATLSSYKTVTAGSESSLESAVN-VNPVSVAIDASHQSOLFYSTSGIYEPECSS 267

292 DPMOVNHAVLAGVGVEDGI-----PYMLIKSWGNWGDNGYFPM 332

268 ENLD-HGVLAGVGGSGSGSSSQSGSSQSGSSNLSASSNEYWIYKXSWGTSWGIEGYILM 325

333 ELGK-NMGCVATCASYPIV 350

326 SRNRDNNGCIASSASFVV 344

RESULT 13

ORVB_ORYSA

ID_ORYSA STANDARD; PRT; 471 AA.

AC P25777;

DT 01-MAY-1992 (Rel. 22, Created)

DI 01-MAY-1992 (Rel. 22, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Oryzain beta chain precursor (EC 3.4.22.-).

OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzeae; Oryza.

OCC NCBI_TaxID=4530;

OCC (1)

RN SEQUENCE FROM N.A.

RC STRAIN=cv. Nipponbare; TISSUE=Seed;

EX MEDLINE=91358494; PubMed=1885617;

RA Watanabe H., Abe K., Emori Y., Hosoyama H., Arai S.;

RT "Molecular cloning and gibberellin-induced expression of multiple

cysteine proteinases of rice seeds (Oryzains).";

J. Biol. Chem. 266:16897-16902(1991).

RL J. Biol. Chem. 266:16897-16902(1991).

CC -!- TISSUE SPECIFICITY: Expressed only in seeds.

CC -!- INDUCTION: By gibberellic acid (GA).

CC -!- SIMILARITY: Belongs to peptidase family C1.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL Outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL; D90407; BAA14403.1; --

PIR; JU0389; KHRZOB.

PDB; IFWO; 09-MAY-01.

MEROPS; C01.029; --

Gnamene; P25777; --

InterPro; IPR000118; Granulin.

InterPro; IPR000668; Peptidase_C1.

InterPro; IPR000169; SHprot_acsite.

Pfam; PF00396; granulin; 1.

Pfam; PF00112; Peptidase_C1; 1.

PRINTS; PR00705; PAPAIN.

ProDom; PD000158; Peptidase_C1; 1.

SMART; SMO0277; GRAN; 1.

SMART; SMO0645; Pept C1; 1.

PROSITE; PS00640; THIOI_PROTEASE_ASN; 1.

PROSITE; PS00139; THIOI_PROTEASE_CYS; 1.

PROSITE; PS00639; THIOI_PROTEASE_HIS; 1.

KW Hydrolase; Thiol protease; Zymogen; Glycoprotein; Signal;

KW 3D-structure.

SIGNAL 1 21 POTENTIAL.

PROPEP 22 139 ACTIVATION PEPTIDE (POTENTIAL).

CHAIN 140 471 ORYZAIN BETA CHAIN.

ACT_SITE 164 164 BY SIMILARITY.

FT ACT SITE 164 164

FT ACT SITE 301 BY SIMILARITY.
FT ACT SITE 321 BY SIMILARITY.
FT DISULFID 161 BY SIMILARITY.
FT DISULFID 195 BY SIMILARITY.
FT DISULFID 295 BY SIMILARITY.
FT CARBOHYD 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 388 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 471 AA; 50505 MW; 0BF39D33995CEBID CRC64;
Query Match 31.5%; Score 587; DB 1; Length 471;
Best Local Similarity 40.2%; Pred. No. 2.9e-39;
Matches 148; Conservative 55; Mismatches 127; Indels 38; Gaps 15;
Qy 2 AAKLLFLLFLV---SALSVALAGFEEDNPIRSYQRPDSIEPAILGLVGLSCRHAFHAF 58
Db 5 AAAAFLLLLVGHRAEDMSIIISNAEHGARGLEEGPTEAE-----ARAAVDLW 54
Qy 59 PARYGKS---YGESEIKKFGIFVENLAFIRSTN-RKDL--SYTLGINQFADLTWEF 112
Db 55 LAENGGS PNALGGEH--RFLVFWNLKFDVAHNARADEGGGFRGLMNRFPADLTNEEF 112
Qy 113 RTNRLGAAQNCATAGHNRHVRDGV--LPVTRDREQIVSPVKDQSCGS-WTFSTTGA 169
Db 113 RATFLGAKVAERSAAGERYHDGVEELPSVDREKGAVPKVKQCGSCWAFSAVST 172
Qy 170 LEAYTQLTGS--TLSSQQLVDCASAFNNFGC--GGLPSQAFYVYKNGGIDTQTPYLG 226
Db 173 VESINQLVTGEMITLSEQLVECTSCNGSCNGLMADAFDFIKNKGIDTDDPYKA 232
Qy 227 VMGICNFKQENGVKVIDSI-NITLGADELKHAAGLVLRPVSVAFEV-VKGENLYKGVY 284
Db 233 VDGKCDINRENAKVSIDGVEDVQNDKSLQKAVA-HQPVSAIEAGGREFQLYHSGVF 291
Qy 285 SSDTCGRDPMVDNHAVALGVGVBDGIPLYLKNKSWGTNNGDNGYFKMELGKNM---CG 340
Db 292 S---GRCGTLSDHGVAVGYGTGDKGDIWYVRNWSGPKGSGYVMERNINVTGKCG 347
Qy 341 VATCASYP 348
Db 348 IAMMASYP 355

RESULT 14
CYSL_HOMAM STANDARD; PRT; 322 AA.
ID_CYS1_HOMAM
AC P13277;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Digestive cysteine proteinase 1 precursor (EC 3.4.22.-).
GN LCPI
OS Homarus americanus (American lobster).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Nephropoidea; Nephropidae; Homarus.
OX NCBI TaxID=6706;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Digestive gland;
RA MEDLINE=92070467; PubMed=1959590;
RX Laycock M.V., Mackay R.M., di Fruscio M., Gallant J.W.;
RA "Molecular cloning of three cDNAs that encode cysteine proteinases in
RT the digestive gland of the American lobster (Homarus americanus).";
RL FEBS Lett. 292:115-120 (1991).
RN [2]
RP SEQUENCE OF 106-133.
RC TISSUE=Digestive juice;
RX MEDLINE=90088376; PubMed=2597115;
RA Laycock M.V., Hirama T., Hasnain S., Watson D., Storer A.C.;
RA "Purification and characterization of a digestive cysteine proteinase
RT from the American lobster (Homarus americanus).";
RL Biochem. J. 263:439-444 (1989).
RN [3]
RP ENZYME REGULATION: Inhibited by E-64, antipain, leupeptin, heavy

CC metal ions, iodacetic acid, dithionitrobenzene, p-hydroxymercuri-
CC benzoate; activated by mercaptoethanol and dithiothreitol.
CC -!- SIMILARITY: Belongs to peptidase family C1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X63567; CA45127.1; -.
DR PIR; S19649; S19649.
DR HSSP; P07711; 1CJL.
DR MEROPS; C01.030; -.
DR InterPro; IPR000668; Peptidase C1.
DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PRO0705; PAPAIN.
DR ProDom; PD000158; Peptidase C1; 1.
DR SMART; SM00645; Pept C1; 1.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
KW Hydrolase; Thiol protease; Zymogen; Signal; Multigene family.
FT SIGNAL 1 16
FT PROPEP 17 105 ACTIVATION PEPTIDE.
FT CHAIN 106 322 DIGESTIVE CYSTEINE PROTEINASE 1.
FT ACT_SITE 129 129 BY SIMILARITY.
FT ACT_SITE 269 269 BY SIMILARITY.
FT ACT_SITE 289 289 BY SIMILARITY.
FT DISULFID 126 170 BY SIMILARITY.
FT DISULFID 160 203 BY SIMILARITY.
FT DISULFID 262 311 BY SIMILARITY.
FT CONFLICT 114 114 K -> E (IN REF. 2).
SQ SEQUENCE 322 AA; 35498 MW; EADCD4E3D31291E9 CRC64;
Query Match 31.4%; Score 586; DB 1; Length 322;
Best Local Similarity 42.3%; Pred. No. 2.2e-39;
Matches 131; Conservative 42; Mismatches 109; Indels 28; Gaps 10;
Qy 59 FARYGKSYGSBEEIKKFGIFVENLAFIRSTNRK----DLSYTLGINQFADLTWEFRT 114
Db 23 FKGFGRKRYVDLEERYRLNVLFDNLQYIEEFNKYERGETYTNLAINQFSDMTNEKFA 82
Qy 115 NRLG-----AAQNCATAGHNRHVRDGVLPVTRDREQIVSPVKDQSCGS-WTFST 166
Db 83 VMKGYKGRPAAVFTSDAPESTEV-----DWRKGAFTPVKQCGSCWAFST 134
Qy 167 TGALEAAVYQLTGS--TLSEQQLVDCASAFNNFGC--GGLPSQAFYVYKNGGIDTQTY 222
Db 135 TGGIEGQHLKTRGLVLSLSEQLVDCAGSYNQCNGGWVERALMYVRDNGSVDTESY 194
Qy 223 PVLGVNGICNFKQENGVKVIDSI-NITLGADELKHAAGLVLRPVSVAFEV-VKGENLYK 281
Db 195 PYEARDNTCRFNSNTIGATCTGYVGIAQGSSEALKTATRDIGPISVAIDASHRSFSY 254
Qy 282 GYVSSDTGCRDPMVDNHAVALGVGVBDGIPLYLKNKSWGTNNGDNGYFKMELGK-NMCG 340
Db 255 GYVFPSCSSQD--HVALGVGSGEGQDFLWKNWSWATSWGSGGYIKMARNNNCG 312
Qy 341 VATCASYPTV 350
Db 313 IATDACYPTV 322
RESULT 15
CATL_DROME
ID CATL_DROME STANDARD; PRT; 341 AA.
AC Q95029; O97431;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

Db	148	CWAFSSTGALSGQHFRKSGVLVLSLSEQLVDCSTKYGNNGCNGGLMDNAPRYIKONGGID	207
Qy	218	TEQTYPLGVMGICNFKQENTGVKVIDSINITLGADELKHAUCLVRPVSVAFEVV-KGF	276
Db	208	TEKSYFYEAIIDDSCHFKNKGTGATDRGFTDIPQGEKKAABAVATVGPVSVAIDASHESF	267
Qy	277	NLYKKGVYSSDTCGRDPMDVNHAFLAVGYGV-EDGIPYWLKNSWGTNMGDNGYFKWELG	335
Db	268	QFYSEGVYNEPQC--DAQNLHDHGLVVGFGTDESGEDYWLVKNSWGTWGDGKGFIKMLRN	325
Qy	336	K-NMCGVATCASYPIV	350
Db	326	KENQCGIASASSYPLV	341

Search completed: March 29, 2004, 22:12:14
Job time : 27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 29, 2004, 22:02:04 ; Search time 79 Seconds

(without alignments)

1405.853 Million cell updates/sec

Title: US-10-087-714-2

Perfect score: 1865
Sequence: 1 MAAKLFFLLFLVLSVAL.....ELGKNMGVATCASYPIVAV 352

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 25:*

- 1: sp_arbhea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	1310.5	70.3	360	10 Q43705	Q43705 zea mays (m
2	1297	69.5	359	10 Q8X178	Q8X178 brassica ol
3	1286	69.0	360	10 Q9LR12	Q9LR12 nicotiana t
4	1282	68.7	359	10 Q9M4E5	Q9M4E5 lolium mult
5	1279.5	68.6	358	10 Q9L7L83	Q9L7L83 arabidopsis
6	1274.5	68.3	358	10 Q8H166	Q8H166 arabidopsis
7	1271	68.2	360	10 Q84YH8	Q84YH8 nicotiana t
8	1265.5	67.9	358	10 Q8WQ9	Q8WQ9 arabidopsis
9	1262.5	67.7	358	10 Q40877	Q40877 petunia hyb
10	1261	67.6	358	10 Q50002	Q50002 prunus arne
11	1254.5	67.3	350	10 Q41057	Q41057 pisum sativ
12	1252.5	67.2	350	10 Q949A2	Q949A2 pisum sativ
13	1207.5	64.7	377	10 Q9W3E7	Q9W3E7 arabidopsis
14	763	40.9	333	11 Q922Q7	Q922Q7 mus musculu
15	763	40.9	335	4 Q96NV6	Q96NV6 homo sapien
16	722	38.7	178	10 Q9SDN1	Q9SDN1 prunus dulc

17	679	36.4	366	5 Q86G38	Q86G38 oxytricha t
18	661.5	35.5	332	5 Q7YW75	Q7YW75 rhipicephal
19	656.5	35.2	326	5 Q24940	Q24940 fasciola he
20	655.5	35.1	326	5 Q3NGW4	Q3NGW4 fasciola gi
21	651.5	34.9	326	5 Q8MUT6	Q8MUT6 fasciola gi
22	648	34.7	335	5 Q7YW74	Q7YW74 rhipicephal
23	647.5	34.7	326	5 Q9XYL8	Q9XYL8 fasciola gi
24	644.5	34.6	326	5 Q24941	Q24941 fasciola he
25	642.5	34.5	326	5 Q721E9	Q721E9 fasciola he
26	641.5	34.4	326	5 Q9NB30	Q9NB30 fasciola he
27	637	34.2	462	10 Q93XQ9	Q93XQ9 ipomoea bat
28	637	34.2	485	10 Q8W180	Q8W180 brassica ol
29	634.5	34.0	326	5 Q9NGW3	Q9NGW3 fasciola gi
30	632.5	33.9	332	5 Q9NHBS	Q9NHBS boophilus m
31	631.5	33.9	311	5 Q9GRW5	Q9GRW5 fasciola he
32	630.5	33.8	310	5 Q96710	Q96710 fasciola he
33	629.5	33.8	311	5 Q8T529	Q8T529 fasciola he
34	627	33.6	325	5 Q26360	Q26360 fasciola sp
35	625.5	33.5	324	5 Q8T0X0	Q8T0X0 fasciola he
36	624.5	33.5	324	5 Q86GJ2	Q86GJ2 hydra atten
37	621.5	33.3	338	5 Q9Y0X2	Q9Y0X2 artemia san
38	621.5	33.3	461	10 Q84M29	Q84M29 helianthus
39	616.5	33.1	351	5 Q8IT42	Q8IT42 theromyzon
40	616	33.0	374	10 Q24137	Q24137 nicotiana t
41	615	33.0	463	10 Q9FWH8	Q9FWH8 arabidopsis
42	613.5	32.9	338	5 Q46030	Q46030 sitophilus
43	612	32.8	374	10 Q84YH7	Q84YH7 nicotiana t
44	611.5	32.8	306	5 Q7Z0G8	Q7Z0G8 metapanaeus
45	611.5	32.8	322	5 Q7Z0G9	Q7Z0G9 metapanaeus

ALIGNMENTS

RESULT 1

Q43705 PRELIMINARY; PRT; 360 AA.

AC Q43705; TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Cysteine protease precursor.
GN SEEL.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_taxid=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. ES; TISSUE=Leaf;
RX MEDLINE=97422404; PubMed=9278172;
RA Griffiths C.M., Hosken S.E., Oliver D., Chojek A.J.S., Thomas H.;
RT "Sequencing, expression pattern and RFLP mapping of a senescence-
enhanced cDNA from Zea mays with high homology to oryzain gamma and
aleurain";
RL Plant Mol. Biol. 34:815-821(1997).
DR EMBL; X99936; CAA68192.1; -;
DR HSP; O46427; 8PCH.
DR GO; GO:0004187; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000668; Peptidase_C1.
DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPAIN.
DR ProDom; PD000158; Peptidase_C1; 1.
DR SMART; SM00645; Pept_C1; 1.
DR PROSITE; PS00640; THIOI_PROTEASE ASN; 1.
DR PROSITE; PS00139; THIOI_PROTEASE_CYS; 1.
DR PROSITE; PS00639; THIOI_PROTEASE_HIS; 1.
DR HydroLase; Protease; Signal; Thiol protease.
FT SIGNAL 29 34
FT CHAIN 143 360
FT POTENTIAL.

Db 13 LLILIAASTASI5GFDESNPIMVSDRLREVEESVVOILQSGSHV5SFAFHAHYKRYE 72

QY 69 SEEEKKFGIFVENLAFIRSTNKKDLSYTLGINQFADLTWEFRNRLGAAQNCATAH 128

Db 73 NAEEMKLRFSIFKENLDLIRSTNKKGLSYKLGYNQFADMTQEFQRTKLGAAQNCSATLK 132

QY 129 GNRHFDVGVLPTWRDREOGIVSPVKDQSCGS-WTFSTTGALAAAYTQLTGS--TLSEQ 185

Db 133 GTHKLTGALPETKDWREDGIVSPVKDQSCGSWTFSTTGALAAAYHQAQFGKISLSQ 192

QY 186 QLVDCASAFNNFGC--GGLPSQAFYIVKINGIDITEQTPYVLGYMGICNFKQENVGVKVD 244

Db 193 QLVDGAGAFNNYCGNGLPSQAFYIKNSGGLDTEAPYTGEDGTCKYSAENVGVVELD 252

QY 245 STNITLGAEDELKHAGLVLRVSVAFVVKNGENLYKKGVYSSTCGRDPMDVNHAVLAVG 304

Db 253 SVNITLGAEDELKHAGLVLRVSVAFVHISFRLYKSGVYSDSHCQGTQMDVNHAVLAVG 312

QY 305 YGVEDGVPYWLIKNSGNTGNWGDNGYFKWELGKMKCVATCASYPIVA 351

Db 313 YGIEDGVYWLIKNSGADWDGKGYFKWEMGKMKGIATCASYPVVA 359

RESULT 3

Q9RLR2 PRELIMINARY; PRT; 360 AA.

ID Q9RLR2 AC Q9RLR2

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)

DE Cysteine protease precursor.

GN NTPC-23

OS Nicotiana tabacum (Common tobacco).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;

OC Lamids; Solanales; Solanaceae; Nicotiana.

CC NCBI_TaxId=4097;

OX [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Samsun NN; TISSUE=Leaf;

RX MEDLINE=21039373; PubMed=11198425;

RA Ueda T, Seo S., Ohashi Y., Hashimoto J.;

RT "Circadian and senescence-enhanced expression of a tobacco cysteine

RT protease gene.;"

RL Plant Mol. Biol. 44:649-657(2000).

RL EMBL; AB032168; BAA96501.1; -.

DR HSSP; P07711; 1CJL.

DR MEROPS; C01.041; -.

DR GO; GO:0004197; P:cysteine-type endopeptidase activity; IEA.

DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

DR InterPro; IPR000569; Peptidase C1.

DR InterPro; IPR000169; SHprot_acsite.

DR Pfam; PF00112; Peptidase_C1; 1.

DR PRINTS; PR00705; PAPAIN.

DR ProDom; PD000158; Peptidase_C1; 1.

DR SMART; SM00645; Pept C1; 1.

DR PROSITE; PS00640; THIOL_PROTEASE ASN; 1.

DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.

DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.

KW Hydrolase; Protease; Signal; Thiol protease.

FT SIGNAL 31 36

FT CHAIN 143 360

FT CYSINE PROTEASE.

SEQUENCE 360 AA; 39199 MW; BFAEC40F2D3CD9B7 CRC64;

Query Match 69.0%; Score 1286; DB 10; Length 360;

Best Local Similarity 69.3%; Pred. No. 7.7e-101;

Matches	246;	Conservative	37;	Mismatches	64;	Indels	8;	Gaps	
QY	5	LLFELLPLVSALSVALAG---	FREDPIRS-VTORPDS	EPAILLVGLVSGRHAFHFAFA	60				
Db	6	LLALLVAGGLFASALAGATFADENPIR	WSDGHELENALLQVGT	TRALLSFAFA	65				
QY	61	PRYKSYGSEBEIKRGRGIVENLAIRSTR	NKDLSTYLLGINQFADLT	WEERTNLGAA	120				

Db 66 HRYGKVESVEIKQRFVFDLNLKMRSHNKKGLSYKLGYNFTDLTWDFRDRLGAA 125
QY 121 QNCSATAGNHRFVDGVLPTVRDREOGIVSPVKDQSCGS-WTFSTTGALEAAAYTOLTG 179
Db 126 QNCSATTGNLKVNVVLPTKDWREAGIVSPVKNQKCGSCWTFTTGALEAAYSQAFG 195
QY 180 S--TLSEQLVDCASAFNNFGC--GGLPSQAFYKYNKGIDTETQTPYLVGMGICNPKQE 236
Db 186 KGISLSEQLVDCAGAFNNFGCGLPSQAFYKYNKGIDTETQTPYLVGMGICNPKQE 245
QY 237 NVGKVIDSINITLGADELKHAVALRVPVSAFVGVKGNLYKGYVSSDTGCRDPMDV 296
Db 246 NVGKVIDSINITLGADELKHAVALRVPVSAFVGVKGNLYKGYVSSDTGCRDPMDV 305
QY 297 NHAVLVGYGVEDGIPWLIKNKSGTNGDNGYFKMELGNMCGVATCASYPIVA 351
Db 306 NHAVLVGYGVENGVPWLIKNKSGADGNGYFKMELGNMCGVATCASYPIVA 360
RESULT 4
Q9M4E5 PRELIMINARY; PRT; 359 AA.
ID Q9M4E5
AC Q9M4E5
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cysteine protease.
GN SEB1.
OS Lolium multiflorum (Italian ryegrass).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Poaceae; Lolium.
OC NCBI_TaxID=4521;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20461783; PubMed=11004495;
RA Li Q., Bettany A.J.E., Donnison I., Griffiths C.M., Thomas H.,
RA Scott I.M.;
RT "Characterisation of a cysteine protease cDNA from Lolium multiflorum
leaves and its expression during senescence and cytokinin treatment.";
RL Biochim. Biophys. Acta 1492:233-236(2000).
DR ENBL; AJ249847; CAB71032.1; -
DR HSP; 046427; 8PCH.
DR MEROPS; C01.041; -
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000668; Peptidase_C1.
DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPAIN.
DR ProDom; PD000158; Peptidase_C1; 1.
DR SMART; SM00645; Pept C1; 1.
DR PROSITE; PS00640; THIOI_PROTEASE ASN; 1.
DR PROSITE; PS00139; THIOI_PROTEASE_CYS; 1.
DR PROSITE; PS00639; THIOI_PROTEASE_HIS; 1.
KW Hydrolase; Protease; Thiol protease.
SQ SEQUENCE 359 AA; 36641 MW; 7595F5FD25C2424 CRC64;
Query Match 68.7%; Score 1282; DB 10; Length 359;
Best Local Similarity 68.6%; Pred. No. 1.7e-100;
Matches 243; Conservative 42; Mismatches 63; Indels 6; Gaps 5;
QY 4 KLLFLLFLVLSALVALAGEENPIRSVTQRDPS-IEPAILVGLSGRHAFHAFARR 62
Db 5 RILVLAVALATAVAASAFDSNPTRPVTERRAASAVESVTLGALGRTHALRFAFVAVR 64
QY 63 YGKSYGSEEBIKRFGIFVFNLAIFSTNRKDI-SYTLGINQFADLTWEERFTRNLGAON 122
Db 65 HGKSYGSAAEVQRRFRIFESLSDEVSTNRKGLSYKLGINRFSMDTWEEFQATLGAAT 124
QY 123 CSATAGNHRFVD-GVLPTVRDREOGIVSPVKDQSCGS-WTFSTTGALEAAAYTOLTGS 180

Db 125 CSATLAGNLMRDANALPETKDWRETGIVSPVKDQSCGSWTFSTTGALEAAAYTQATGK 184
QY 181 --TLSEQLVDCASAFNNFGC--GGLPSQAFYKYNKGIDTETQTPYLVGMGICNPKQEN 237
Db 185 NLSLSEQLVDCAGAYNFCNGGLPSQAFYKYNKGIDTETQTPYLVGMGICNPKQEN 244
QY 238 VGKVIDSINITLGADELKHAVALRVPVSAFVGVKGNLYKGYVSSDTGCRDPMDV 297
Db 245 AAQVQVADSVNITLNADELKHAVALRVPVSAFVGVKGNLYKGYVSSDTGCRDPMDV 304
QY 298 HAVLVGYGVEDGIPWLIKNKSGTNGDNGYFKMELGNMCGVATCASYPIVA 351
Db 305 HAVLVGYGVENGVPWLIKNKSGADGNGYFKMELGNMCGVATCASYPIVA 358
RESULT 5
Q9LL83 PRELIMINARY; PRT; 358 AA.
ID Q9LL83
AC Q9LL83
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cysteine proteinase AALP (Putative AALP protein) (Hypothetical protein).
GN AT5G60360.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsia.
OC NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20330156; PubMed=10871276;
RA Ahmed S.U., Rojo E., Kovaleva V., Venkataraman S., Dombrowski J.E.,
RA Matsuoka K., Raikhel N.V.;
RT "The plant vacuolar sorting receptor AtELP is involved in transport of
NH(2)-terminal propeptide-containing vacuolar proteins in Arabidopsis
thaliana.";
RL J. Cell Biol. 149:1335-1344(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=98344145; PubMed=9679202;
RA Kaneko T., Kotani H., Nakamura Y., Sato S., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. V. Sequence
features of the regions of 1,381,565 bp covered by twenty one
physically assigned P1 and TAC clones.";
RL DNA Res. 5:131-145(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Full length cDNA of gene MUF9.1/AT5G60360 (GI:9757740).";
RN Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Haas B.J., Volkovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
annotation.";
RL Genome Biol. 0:0-0(2002).
RN [5]
RP SEQUENCE FROM N.A.
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana.";
RN Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

DT	01-MAR-2003 (TrEMBLrel. 23, Created)
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Putative cysteine proteinase AAPF.
GN	ATSG60360.
OS	Arabisopsis thaliana (Mouse-ear cress).
OC	Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; rosids;
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;	
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
RA	Quach H.L., Tang C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
RA	Yu G., Yuan S., Carrinzi P., Chen H., Cheuk R., Hayashizaki Y.,
RA	Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,
RA	Nuyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
RA	Southwick A., Tripp M.G., Wu T., Shinzaki K., Davis R.W., Ecker J.R.,
RA	Theologis A.; Full Length cDNA Clones."
RL	Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR	ENML: EF000674; RAN1820.1; "
DR	GO: GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR	GO: GO:0008508; P:proteolysis and peptidolysis; IEA.
DR	InterPro: IPR000668; Peptidase_C1.
DR	InterPro: IPR000169; SHprot_acsite.
DR	Pfam: PF00112; Peptidase_C1; 1.
DR	PRINTS: PR00705; PAPAIN.
DR	ProDom: PD000158; Peptidase_C1; 1.
DR	SMART: SM00645; Pept_C1; 1.
DR	PROSITE: PS00640; THIOLE PROTEASE ASN; 1.
DR	PROSITE: PS00139; THIOLE PROTEASE_CYS; 1.
DR	PROSITE: PS00639; THIOLE PROTEASE_HIS; 1.
SQ	SEQUENCE 358 AA; 39007 MW; 2524405745617F17 CRC64;

Query Match 68.3%; Score 1274.5; DB 10; Length 358;
 Best Local Similarity 66.8%; Pred. No. 7.2e-100;
 Matches 239; Conservative 39; Mismatches 73; Indels 7; Gaps 4

QY	1	MAAKLLF--FLFLVLSALVALAGFEEDPIRSVTQRPSIRPAILGVLGSCRHAFHPA	57
Db	1	MSAKTILSSVLVLVLAASAAANGFDSPRIWSDGLREVESVSQILGQSRHVLSFA	60
QY	58	RRARYGKSYGBEIKRFGIFVENLAFTSNRKDLSTYLGINOFADLTWEERFTNEL	117
Db	61	RTRHYGKKYNQVSEMDKRFIFKENLDLRSTNKKGSLYKLGVNFADLTWQEFOETKL	120
QY	118	GAAQNCSATAHGNHRFDVGLVPVTRDWRBEGIVSPVKDGSCGS-WTFSTTGALAAAYTQ	176
Db	121	GAAQNCSATLKSHKHVTEAALPETKWREGIVSPVKDGCGCSGWTFSTTGALAAAYHQ	180
QY	177	LRTGS--TLSEEQLVDCSAFNFGC-GGLPSQAPEFYKVGNGGIDTPTQTPYPLGVGMICNF	233
Db	181	AFPGKISLSEQLVDCAGAFNNYCNGSLFSQAPEFYKSNGLDTEKAYPYTKDGTCKP	240
QY	234	KQENVGVKVDSINITLGADELGHAVGLVRPVSVAFEVKGNLYKKYGVSDDTCGRP	293
Db	241	SAENVGVQLVNSVNITLGADELGHAVGLVRPVSVIAFEVILHSFRLYKGYVDTHSGGSTP	300
QY	294	MDVNHAVLAGVGVEDGIPYWLINSGTNWDNGYFKOMELGRMCGVATCASYPIVA	351
Db	301	MDVNHAVLAGVGVEDGPVYWLKNKGADWGDKGYFKOMEMGNMGCIATCASYPVA	358

RESULT 7
 Q84YH8 PRELIMINARY; PRT; 360 AA.

ID	Q84YH8
AC	Q84YH8;
DT	01-JUN-2003 (TrEMBLrel. 24, Created)
DT	01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	NTPC23-like cysteine proteinase.
OS	Nicotiana tabacum (Common tobacco)

DT 01-JUN-2003 (TREMBLrel. 24, created)
DT 01-JUN-2003 (TREMBLrel. 24, last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, last annotation update)
DE NTPC23-like cysteine proteinase.
OS Nicotiana tabacum (Common tobacco).


```

DR InterPro: IPR000169; SHprot.acssite.
DR Pfam: PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPAIN_
DR ProDom; PD000158; Peptidase_C1; 1.
DR SMART; SM00645; Rept_C1; 1_
DR PROSITE; PS00038; HLH 1; 1.
DR PROSITE; PS00640; THIOL_PROTEASE ASN; 1.
DR PROSITE; PS00139; THIOL_PROTEASE CYS; 1.
DR PROSITE; PS00639; THIOL_PROTEASE HIS; 1.
KW Hydrolase; Protease; Thiol protease.
SQ SEQUENCE 358 AA; 39309 NM; C98F78793B002554 CRC64;

      Query Match          67.6%; Score 1261; DB 10; Length 358;
      Best Local Similarity 68.0%; Pred. No.1e-98;
      Matches 236; Conservative 35; Mismatches 72; Indels 4; Gaps 3

QY   9  LLPLVSLVALGAFFEDNPIRSTQPDIEPAILGLGSCRHAFHFAFRRYCKSVG 68
Db   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY  69  SEEEKRGFGIFVENLAFIRSTNKKDLSTYLGINQFADLTWEFRNRLGAQNCSAT 128
Db   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY  72  SYEENKLAYEIFSENKKLIRSTNKKGLPYTLAVNRFADWSWEFRQRRLGAQNCSAT 131
Db   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY  129  GNHRFVDGVLPTDWRREQVTSVPKDGSCGS-WPFSTGALBAAYTQLTGS--TL 185
Db   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY  132  GSREHTDAVLPSKKNWREGILVTPKQGHGCGSWFTTGAUEAAYVAQFRKQSLSE 191
Db   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY  186  QLVDCASAFNFGC--GGLPSQAFEPVYKNGIGDTEQTPYGLVMGICFNFKQENVGV 244
Db   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY  192  QLVDQACAGAFNFGCHGGLPSQAFEYIKYNGSLDTEAAYPVYGDGACKFSAENV 251
Db   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY  245  SLNITLGAEDELKHAUGLVRPVSAFRRVWKGFNLYKKGVVSSDTCGRDPMDNH 304
Db   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY  252  SVNITLGDQEQLKHAFAFVPVSAFQVVKSAFRIYKSGVYTSDTCCSSPMDN 311
Db   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY  305  YGVEDGIPYFWLIKNSWGTNMGDNGYFKNELGKNMCGVATCASYPIVA 351
Db   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY  312  YVEEGGVFWLIKNSWGESGNDNGYFKNELGKNMCGVATCASYPIVA 358

```

RESULT 11	
Q41057	PRELIMINARY; PRT; 350 AA.
ID	
AC	Q41057;
DT	01-NOV-1996 (TrEMBLrel. 01, Created)
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE	Cysteine protease
OS	Pisum sativum (Garden pea)
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC	eurosid 1; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
OX	NCBI_TaxID=3888;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=Peilham First;
RA	MEDLINE=96350414; PubMed=8765523;
RX	Jones C.G., Tucker G.A., Lycett G.W.;
RT	"Pattern of expression and characteristics of a cysteine protease cDNA
RL	from pea (Pisum sativum L.).";
RL	Biochim. Biophys. Acta 1296:13-15(1996).
DR	EMBL; Z68291; CAA92593.1; -.
DR	PIR; S71923; S71923.
DR	HSSP; O46427; 8PCH.
DR	MEROPS; C01.041; -.
DR	GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR	InterPro; IPR000668; Peptidase_C1.
DR	InterPro; IPR000169; SHprot_acsSite.
DR	Pfam; PF00112; Peptidase_C1; 1.
DR	PRINTS; PRO0705; PAPAIN.
DR	ProDom; PD000158; Peptidase C1; 1.

DR SMART; SM00645; Pept C1; 1.
 DR PROSITE; PS00640; THIOLESTERASE ASN; 1.
 DR PROSITE; PS00139; THIOLESTERASE_CYS; 1.
 DR PROSITE; PS00639; THIOLESTERASE_HIS; 1.
 KW Hydrolyase; Protease; Thiol protease.
 SQ SEQUENCE 350 AA; 38575 MW; 0D7F8F2C51531893 CRC64;

Query Match 67.3%; Score 1254.5; DB 10; Length 350;
 Best Local Similarity 67.2%; Pred. No. 3.5e-98;
 Matches 240; Conservative 39; Mismatches 65; Indels 13; Gaps 6;

QY 1 MAAKLLFFLLFLYSALVALAG--FEEDNPIRSVTORPDSIEPAILGLVLSCHRAHFAH 58
 DB 1 MAQWLSLLIVLCFVAS---AAAGFSFHDSPNIRMVSV---DVEEQLLQVIGESHAHVSFA 53
 QY 59 FARRYGKSGSEEEIKKRGIFVENLAFIRSTNRKDLSTYLGINQFADLTWEEFRNRLG 118
 DB 54 FANRYGKGYSDVDKMLFKIFSENLELRSSNKRRLSYKLGVNHFPADMTWEEFRSHRLG 113
 QY 119 AAQNCSTAHGHRFVDGVLPTVRDMREOGIIVSVKDGSCGS-WTFSTTGALSAAYTQL 177
 DB 114 AAQNCSTATLGNHKTIDANLPDKWRKEGIVSGVKDQSCGSCWTFSTTGALSAAYA 173
 QY 178 TGS--TLSEQQQLVDCASAFNNFGC--GGLPSQAFYVYKNGGIDTEQTPYLVGMGICNFK 234
 DB 174 FGKNTLSLSEQQQLVDCAGAFNNFGCGSLPSQAFYIKYNGGLETEEAYPTVTSGLCKFR 233
 QY 235 QENVGVKVIDSINITLGADELKHAIVLRPVSAFVYKNGVFNLYKKGVSSTDCGRDPM 294
 DB 234 SEHVAVKVLSGWNITLGADELKHAIAFARPVSAFVYVVDHFRLYKSGVYTSTACGSTM 293
 QY 295 DVNHAVLAGYGVDEGIPYWLKNSGNTWNGDGYFKMELGKMGKMGVATCSYPIVA 351
 DB 294 DVNHAVLAGYGVDEGIPYWLKNSGNTWNGDGYFKMELGKMGKMGVATCSYPIVA 350

RESULT 12
 Q949A2 ID Q949A2 PRELIMINARY; PRT; 350 AA.
 AC Q949A2;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Early leaf senescence abundant cysteine protease.
 GN ELSA.
 OS Pisum sativum (Garden pea).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
 OX NCBI_TaxID=3888;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Messire;
 RA Pic E.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ278699; CAC41636.1; -.
 DR MEROPS; C01.041; -.
 DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000668; Peptidase_C1.
 DR InterPro; IPR000169; SHprot_acsite.
 DR Pfam; PF00112; Peptidase_C1.
 DR PRINTS; PR00705; PAPAIA.
 DR SMART; SM00645; Pept C1; 1.
 DR PROSITE; PS00640; THIOLESTERASE ASN; 1.
 DR PROSITE; PS00139; THIOLESTERASE_CYS; 1.
 DR PROSITE; PS00639; THIOLESTERASE_HIS; 1.
 KW Hydrolyase; Protease; Thiol protease.
 SQ SEQUENCE 350 AA; 38575 MW; 0CBFF2F87B61892 CRC64;

Query Match 67.2%; Score 1252.5; DB 10; Length 350;
 Best Local Similarity 66.9%; Pred. No. 5.2e-98;

Matches 239; Conservative 40; Mismatches 65; Indels 13; Gaps 6;

QY 1 MAAKLLFFLLFLYSALVALAG--FEEDNPIRSVTORPDSIEPAILGLVLSCHRAHFAH 58
 DB 1 MAQWLSLLIVLCFVAS---AAAGFSFHDSPNIRMVSV---DVEEQLLQVIGESHAHVSFA 53
 QY 59 FARRYGKSGSEEEIKKRGIFVENLAFIRSTNRKDLSTYLGINQFADLTWEEFRNRLG 118
 DB 54 FANRYGKGYSDVDKMLFKIFSENLELRSSNKRRLSYKLGVNHFPADMTWEEFRSHRLG 113
 QY 119 AAQNCSTAHGHRFVDGVLPTVRDMREOGIIVSVKDGSCGS-WTFSTTGALSAAYTQL 177
 DB 114 AAQNCSTATLGNHKTIDANLPDKWRKEGIVSGVKDQSCGSCWTFSTTGALSAAYA 173
 QY 178 TGS--TLSEQQQLVDCASAFNNFGC--GGLPSQAFYVYKNGGIDTEQTPYLVGMGICNFK 234
 DB 174 FGKNTLSLSEQQQLVDCAGAFNNFGCGSLPSQAFYIKYNGGLETEEAYPTVTSGLCKFR 233
 QY 235 QENVGVKVIDSINITLGADELKHAIVLRPVSAFVYKNGVFNLYKKGVSSTDCGRDPM 294
 DB 234 SEHVAVKVLSGWNITLGADELKHAIAFARPVSAFVYVVDHFRLYKSGVYTSTACGSTM 293
 QY 295 DVNHAVLAGYGVDEGIPYWLKNSGNTWNGDGYFKMELGKMGKMGVATCSYPIVA 351
 DB 294 DVNHAVLAGYGVDEGIPYWLKNSGNTWNGDGYFKMELGKMGKMGVATCSYPIVA 350

RESULT 13
 Q9M3E7 ID Q9M3E7 PRELIMINARY; PRT; 377 AA.
 AC Q9M3E7;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Cysteine protease-like protein.
 GN F18N11.70.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopses.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Jordan N., Bangert S., Wiedemann R., Voss H., Unsel M., Mewes H.W.,
 Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.
 RA Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL132953; CAB72480.1; -.
 DR PIR; T47471; T47471.
 DR HSP; O46427; SPCH.
 DR MEROPS; C01.041; -.
 DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000668; Peptidase_C1.
 DR InterPro; IPR000169; SHprot_acsite.
 DR Pfam; PF00112; Peptidase_C1.
 DR PRINTS; PR00705; PAPAIA.
 DR PRODOM; PD000158; Peptidase_C1; 1.
 DR SMART; SM00645; Pept C1; 1.
 DR PROSITE; PS00640; THIOLESTERASE ASN; 1.
 DR PROSITE; PS00139; THIOLESTERASE_CYS; 1.
 DR PROSITE; PS00639; THIOLESTERASE_HIS; 1.
 KW Hydrolyase; Protease; Thiol protease.
 SQ SEQUENCE 377 AA; 41835 MW; DDC082A6739A180 CRC64;

Query Match 64.7%; Score 1207.5; DB 10; Length 377;
 Best Local Similarity 65.9%; Pred. No. 3.8e-94;
 Matches 228; Conservative 37; Mismatches 74; Indels 7; Gaps 4;

QY 1 MAAKLLP---FULLFLYSALVALAGFEEDNPIRSVTORPDSIEPAILGLVLSCHRAHFAH 57

```

Db 1 MSVKNLSSILLTLPAAASKEIGFDESFPKKNWSDNHELEDIVVQILGOSRHVSFS 60
QY 58 RPARYKSGYSGEBEIKRFGIFVENLAFIRSTNRKDLSTYLGINQFADLTWEEFR 117
Db 61 RETHYKGYKYSVEEMKLRFSFKENLDLIRSNKGLSYKLSNQFADLTWQEFQYKL 120
QY 118 GAAQNCSTATAHNRHFDVGLPVTDRWREOGIVSPVKDQSGS-WTSTTTCALAAAYTQ 176
Db 121 GAAQNCSTATLKGSHKITEATVPTDKWREDGIIVSPVKEQCHGCGSWTSTTTCALAAAHQ 180
QY 177 LTGS--TLSEQLVDCASAFNFGC--GGLPSQAFYKYNKGGIDTEQTYPLGVGICNF 233
Db 181 AFGKISLSEQLVDCAGTFNFGCHGLFSQAFYKYNKGGIDTEQTYPLGVGICNF 240
QY 234 KOENGVKVIDSINITIGAEDELKHAVALRVPVSAFVVKGNLYKKGVSSTDCGRDP 293
Db 241 SAKNTGVQVRSVNITLGADELKHAVALRVPVSAFVVKGNLYKKGVSSTDCGRDP 300
QY 294 MDVNHAVLAGVGVEDGIPYWLKNSGNTNGDNGYFKMELGNKMC 339
Db 301 MDVNHAVLAGVGVEDDIPYWLKNSGNTNGDNGYFKMELGNKMC 346

RESULT 14
Q922Q7 PRELIMINARY; PRT; 333 AA.
ID Q922Q7 AC Q922Q7
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Cathepsin H.
GN CTSH
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BCC06878; AAH06878.1; -.
DR MGD; MGI:107285; Ctab.
DR GO; GO:0004197; P:cytosine-type endopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPAIN.
DR ProDom; PD000158; Peptidase_C1; 1.
DR SMART; SM00645; Pept C1; 1.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KW Hydrolase; Protease; Thiol protease.
SQ SEQUENCE 333 AA; 37155 MW; C28B344AC3439DE8 CRC64;

```

```

Query Match 40.9%; Score 763; DB 11; Length 333;
Best Local Similarity 50.8%; Pred. No. 1.9e-56;
Matches 155; Conservative 49; Mismatches 93; Indels 8; Gaps 7;

QY 54 FHFAFARYKSGYSGEBEIKRFGIFVENLAFIRSTNRKDLSTYLGINQFADLTWEEFR 113
Db 31 FHFKSWMKQKTY--SSVEYNHRLQMPANNRKLQAHNQNRNHTFKNALNQSDMSFABIK 89
QY 114 TNRL-GAQNCSATAGHNRHFDVGLPVTDRWREOG-IVSPVKDQSGS-WTSTTTCAL 170
Db 90 HFLWSEFQNCAT-KSNYLRGTGYPYSDMRKKNVSPVINGAGCSWTSTTTCAL 148
QY 171 EAAATQLTGSTLS--EQQLVDCASAFNFGC--GGLPSQAFYKYNKGGIDTEQTYPLGV 227
Db 149 ESAVAIAGKMLSLAEQQLVDCAFNNGCKGLPSQAFYKYNKGGIDTEQTYPLGV 208

```

```

QY 228 MGICNFQKQENGVKVIDSINITIGAEDELKHAVALRVPVSAFVVKGNLYKKGVSST 287
Db 209 DSSCREFPQKAVAFVKVNVNITLNDBAAMVEAVALNPNVSAFVTEDFLTKSGVYSK 268
QY 288 TCGRDPMDVNHAVLAGVGVEDGIPYWLKNSGNTNGDNGYFKMELGNKMCVATCASY 347
Db 269 SCHKTPDKVNHAVLAGVGVEDGIPYWLKNSGNTNGDNGYFKMELGNKMCVATCASY 328
QY 348 PIVAV 352
Db 329 PIPQV 333

RESULT 15
Q96NY6 PRELIMINARY; PRT; 335 AA.
ID Q96NY6 AC Q96NY6
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Cathepsin H.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Waghray A., Keppeler D., Sloane B.F., Chen Y.Q.;
RL "Analysis of a truncated form of cathepsin H in human prostate tumor cells.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF426247; AAL23961.1; -.
DR GO; GO:0004197; P:cytosine-type endopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000668; Peptidase_C1.
DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF00112; Peptidase_C1; 1.
DR PRINTS; PR00705; PAPAIN.
DR ProDom; PD000158; Peptidase_C1; 1.
DR SMART; SM00645; Pept C1; 1.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KW Hydrolase; Protease; Thiol protease.
SQ SEQUENCE 335 AA; 37419 MW; B37P214450E26170 CRC64;

Query Match 40.9%; Score 763; DB 4; Length 335;
Best Local Similarity 52.1%; Pred. No. 1.9e-56;
Matches 159; Conservative 40; Mismatches 98; Indels 8; Gaps 7;

QY 54 FHFAFARYKSGYSGEBEIKRFGIFVENLAFIRSTNRKDLSTYLGINQFADLTWEEFR 113
Db 33 FHFKSWMKQKTY-STEEYHHELOTFASNRKKNHNNHTFKNALNQSDMSFABIK 91
QY 114 TNRL-GAQNCSATAGHNRHFDVGLPVTDRWREOG-IVSPVKDQSGS-WTSTTTCAL 170
Db 92 HYLWSEFQNCAT-KSNYLRGTGYPYSDMRKKNVSPVINGAGCSWTSTTTCAL 150
QY 171 EAAATQLTGSTLS--EQQLVDCASAFNFGC--GGLPSQAFYKYNKGGIDTEQTYPLGV 227
Db 151 ESAIAIATGKMLSLAEQQLVDCAFNNGCKGLPSQAFYKYNKGGIDTEQTYPLGV 210
QY 228 MGICNFQKQENGVKVIDSINITIGAEDELKHAVALRVPVSAFVVKGNLYKKGVSST 287
Db 211 DGYCKFGKALGFVKDVAITTYDEAMVEAVALNPNVSAFVTEDFLTKSGVYSK 270
QY 288 TCGRDPMDVNHAVLAGVGVEDGIPYWLKNSGNTNGDNGYFKMELGNKMCVATCASY 347
Db 271 SCHKTPDKVNHAVLAGVGVEDGIPYWLKNSGNTNGDNGYFKMELGNKMCVATCASY 330
QY 348 PIVAV 352
Db 331 PIPV 335

```

Search completed: March 29, 2004, 22:15:12
Job time : 81 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 29, 2004, 22:22:05 ; Search time 2371 Seconds
(without alignments)
4433.356 Million cell updates/sec

Title: US-10-087-714-2
Perfect score: 1865
Sequence: 1 MAAKLFFLLFLVSALSVAL.....ELGKMGVATCASYPIVAV 352

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Ddelop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US10087714/runat 26032004 113453 946/app query.fasta 1.519
-DB=EST -QPMT=fastap -SUFFIX=rest -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10087714 -CGN 1 1 2135 @runat 26032004 113453 946 -NCPU=3
-NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -PGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1321.5	70.9	1532	11	AY106223 Zea mays
2	1097.5	58.8	989	14	CK263045
3	1038	55.7	985	14	CK287181
4	1035	55.5	829	14	CB341855
5	1034	55.4	977	14	CK295860
6	1022.5	54.8	1005	14	CF212812
7	1014.5	54.4	936	14	CK268548
8	1004.5	53.9	920	14	CK257583
9	988	53.0	748	14	CB982399
10	987	52.9	814	12	BG594735
11	981	52.6	965	14	CK294622
12	975.5	52.3	731	13	BQ281381
13	970	52.0	768	14	CB892658
14	967.5	51.9	794	10	BE19255
15	957	51.3	720	10	BE131652
16	944	50.6	924	14	CK294694
17	943	50.6	930	14	CK295057
18	941	50.5	946	14	CK285936
19	938	50.3	848	14	CF510505
20	934	50.1	886	14	CF510584
21	925	49.6	901	14	CK298522
22	919.5	49.3	910	13	BQ797116
23	919	49.3	701	14	CB894684
24	918.5	49.2	1230	10	BE422216
25	915.5	49.1	852	14	CK263154
26	915	49.1	760	10	BF267048
27	915	49.1	888	14	CK292918
28	915	49.1	911	14	CK285268
29	912	48.9	851	14	CB891143
30	910	48.8	653	14	CB917819
31	910	48.8	899	14	CK293298
32	909.5	48.8	709	12	BG155091
33	901	48.3	755	12	BG645271
34	891	47.8	894	13	BU039998
35	885	47.5	703	12	BE263594
36	884.5	47.4	827	12	BG645454
37	883.5	47.4	686	14	CF477066
38	880.5	47.2	842	14	CF517076
39	880	47.2	648	13	BQ625193
40	879.5	47.2	844	14	CF518186
41	876	47.0	755	14	CB981790
42	875	46.9	635	13	BU039803
43	874.5	46.9	639	10	AW158820
44	874.5	46.9	665	10	AW399964
45	874	46.9	710	12	B1267420

ALIGNMENTS

RESULT 1
AY106223
LOCUS Zea mays PC0114759 mRNA sequence.
DEFINITION Zea mays
ACCESSION AY106223
VERSION AY106223.1 GI:21209301
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 1532)

/tissue type="abiotic stress treated leaf and root tissue"
 /lab host="DH10B-Tona"
 /clone_lib="potato abiotic stress cDNA library"
 /notes="vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;
 supplier: Solum tuberosum var. Kennebec plants were
 grown from cuttings on a 16hr light/8 hr dark cycle at 25
 C for 3-4 weeks. Abiotic stress conditions were applied to
 four separate sets of plants. Set 1 involved saturation of
 the soil with 150 mM NaCl and tissues were harvested at
 following application of the salt stress (leaves: 2hr,
 6hr, 12hr, 1d, 2d, and 4d; roots: 2hr, 6hr, 12hr, and 2d).
 Set 2 were grown under the standard conditions and then
 were water stressed by withdrawal of further watering
 applications. Drought stressed plants were harvested after
 cessation of watering (leaves: 3d, 5d, and 7d; roots: 3d
 and 5d). Set 3 were grown under the standard conditions
 and then were cold stressed by placement at 4 C. Cold
 stressed leaves were harvested at 2 hr, 6 hr, 12 hr, 1 d,
 and 4d and roots were harvested at 2 hr, 6 hr, 12 hr, 1 d,
 2d. Set 4 were grown under the standard conditions and
 then were heat stressed by placement at 35 C. Heat
 stressed leaves were harvested at 2 hr, 6 hr, 12 hr, 1 d,
 2d and 4d and heat-stressed roots were harvested at 6 hr,
 12 hr, 1 d, and 4d. RNA was isolated from all tissues and
 equal RNA from each tissue and stress was pooled to
 construct the cDNA library. RNA sample."

ORIGIN

Alignment Scores:
 Pred. No.: 5,76e-120 Length: 989
 Score: 1097.50 Matches: 220
 Percent Similarity: 77.30% Conservative: 32
 Best Local Similarity: 67.48% Mismatches: 65
 Query Match: 58.85% Indels: 9
 DB: 14 Gaps: 6

US-10-087-714-2 (1-352) x CK263045 (1-989)

Qy 6 LeupheLeuLeuPheLeuValSerAlaLeu---SerValAlaLeuAlaGly----- 22
 Db 15 CTCTCACTCTATTGGTTCGTCGCGCGCCTTTTCTGCTGCTCACTTGC CGGACCGCG 74
 Qy 23 ---PheGluGluAspAsnProIleArgSerValThrGlnArgProAspSerIleGluPro 41
 Db 75 ACTTTCGCCGATGAGATCCGATCGACGAGTGTAGT---TCCGAGAGCTGGAGAAC 131
 Qy 42 AlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAlaArg 61
 Db 132 GGAATTCCTCAAGTCGTCGCCAGACTCGCAATGCTCTCTCTGCTGCTTTGCTATC 191
 Qy 62 ArgTyrGlyLysSerTyrGlySerGluGluGluIleLysLysArgPheGlyIlePheVal 81
 Db 192 AGGCATCGGAAGAGTACGAGTCGTTGAGGAGATCAAGCAAGGTTCCGAGATATTTTG 251
 Qy 82 GluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyrThrLeuGlyIle 101
 Db 252 GACAATCTGAAGATGATCCGATCGCATAACAGCAAGAGGACTATACAACTCGGTGC 311
 Qy 102 AsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAlaAlaGln 121
 Db 312 AATGAGTTTACCCACCTAATACATGGATGAGTTCCTAGACCAAGGTTGGGGCATCTCAA 371
 Qy 122 AsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProValThr 141
 Db 372 AACTGTTCTGCCACTACAAGGCGATCTCAAGCTAACTAACGTTGTTTCCGAGAGCG 431
 Qy 142 ArgAspTrpArgGluGlnGlyIleValSerProValLysAspGlnGlySerCysGlySer 161
 Db 432 AAGGACTCGAGGAAGATGATTGTTAGCCAGTGAAGGACAGGCGCAAGTGGCGATCT 491
 Qy 162 ---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGlySer 180
 Db 492 TGTGACATTCAGCACTCTGTGTCATAGAGGCGAGCATATGCCCAAGCATTTGGGAAG 551

Qy 181 -----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnPhenGly 198
 Db 552 GGAATCTCTCTCTCAGAGCAGCAGCTTGTGGACTGTGCTGGAGCTTTTAATAACTTTGGC 611
 Qy 199 Cys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGlyLeuAsp 217
 Db 612 TCAATGGGGGGTGGCCATCACAGCCTTGGTACATTAATTAATTCATGGTGTCTTGAC 671
 Qy 218 ThrGluGlnThrTyrProTyrLeuGlyValMetGlyIleCysAsnPhenLysGlnGluAsn 237
 Db 672 ACTGAAGAGCATATCCATACACCGCAAGATGCATATGTAATTTCTCAACAGCAAT 731
 Qy 238 ValGlyValLysValIleAspSerIleAsnIleThrLeuGlyValAlaGluAspGluLeuLys 257
 Db 732 ATGGTGTCAAGATCATCGTTCTGTCAATATTACCTGGGTGGTGAAGATGAATGAA 791
 Qy 258 HisAlaValGlyLeuValArgProValSerValAlaPheGluValValLysGlyPheAsn 277
 Db 792 TAGGCAGTTGCTAGTGGTCTAGTGTCTTTAGGTGGTAAAGTTTCAA 851
 Qy 278 LeuTyrLysLysGlyValTyrSerSerAspThrCysGlyArgAspProMetAspValAsn 297
 Db 852 CAGTATAAGAGCGGAGTTTACACGACACTGAATGTGGCAGACTCCCATGGACGTAAC 911
 Qy 298 HisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIleProTyrTrpLeuIleLys 317
 Db 912 CATGCTGTTCTTGTGTTGGGTACGTTTGAATGGGGTCCCTATCTGCTCATAAAG 971
 Qy 318 AsnSerTrpGlyThrAsn 323
 Db 972 AACTCATGGGAGCAGAT 989
 RESULT 3
 LOCUS CK287181
 DEFINITION ES7749903 Nicotiana benthamiana mixed tissue cDNA library,
 normalized, full-length Nicotiana benthamiana cDNA clone NEMB738 5',
 end, mRNA sequence.
 ACCESSION CK287181
 VERSION CK287181
 KEYWORDS CK287181.1 GI:39863480
 SOURCE Nicotiana benthamiana
 ORGANISM Nicotiana benthamiana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
 asterids; lamiales; Solanales; Solanaceae; Nicotiana.
 REFERENCE 1 (bases 1 to 985)
 AUTHORS Buell C.R., Hart A., Zismann V., Karamycheva, S.A., Day, B.,
 Staskawicz, B., Jin, H. and Baker, B.
 TITLE Generation of EST sequences from Nicotiana benthamiana
 JOURNAL Unpublished (2003)
 COMMENT Other ESTs: ES7749904
 Contact: Robin Buell
 The Institute for Genomic Research
 9712 Medical Center Dr. Rockville, MD 20850, USA
 Email: potato-array@tigr.org
 Clones can be requested from TIGR via potato@tigr.org
 Seq primer: ATT TAG GTG ACA CTA TAG.
 Location/Qualifiers
 1. 985
 /organism="Nicotiana benthamiana"
 /mol_type="mRNA"
 /db_xref="taxon:4100"
 /clone="NEMB738"
 /tissue_type="abiotic and biotic stress-treated leaves,
 callus tissue and root tissue"
 /lab host="DH10B-Tona"
 /clone_lib="Nicotiana benthamiana mixed tissue cDNA
 library, normalized, full-length"
 /note="vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;
 supplier: RNA was isolated from Nicotiana benthamiana
 tissues that include callus, roots from liquid culture


```

DB: 14 Gaps: 3
US-10-087-714-2 (1-352) x CB341855 (1-829)
QY 42 AlalileLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAlaArg 61
Db 2 TCCGTCCTCCGCTCGCTGCGGACACGCGTACGCTACTCTCTTCGCGAGCTTCGCTCAC 61
QY 62 ArgTyrGlySerTyrGlySerGluGluGluLeuLysArgPheGlyLeuPheVal 81
Db 62 AGGTATGGGAAGAGTTACAAGCGGTGGACGAGATTAACTGAGTTCGAGATTTCTTCG 121
QY 82 GluAsnLeuAlaPheAlaArgSerThrAsnArgLysAspLeuSerTyrThrLeuGlyLe 101
Db 122 GAGAAATTTGAACTCATCAGATCCACCAAGAGGCTTGCCTTATCTACTCTAGCTGT 181
QY 102 AsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyValAlaAlaGln 121
Db 182 AATCAGTTCGCTGATGACCTGGAGAGATTCCGCAGACACAGGTTGGAGCTGCTCAG 241
QY 122 AsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProValThr 141
Db 242 AACTGCTCTGCACCTTGAGGGCAATCACAAGCTAACTGACGTATCTCTCTGAGACG 301
QY 142 ArgAspTrpArgGluGluGlnGlyLeuValSerProValLysAspGlnGlySerCysGlySer 161
Db 302 AAAGATTGAGAGAGATGGCATAGTCAGCCCAATCAAGATCAAGTCACTGTGGATCT 361
QY 162 ---TrpThrPheSerThrThrGlyValAlaLeuGluAlaAlaTyrThrGlnLeuThrGlySer 180
Db 362 TGCTGGACTTTTCAGCACCTGGAGCTCTAGAGGCAGCTTACGCTCAGGCATTTGGGAG 421
QY 181 -----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPheGly 199
Db 422 GGGATCTCTCTGTGAGCAGCAGCTGTGTGACTGTGCGGAGCTTTCAATAAATTTTGA 481
QY 199 Cys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGlyLeuAsp 217
Db 482 TGCACGGGGATTGCCATCCCAAGCTTTTGAGTATACATAAATACATGTTGGCTTGAT 541
QY 218 ThrGluGlnThrTyrProTyrLeuGlyValMetGlyLeuCysAsnPheLysGlnGluAsn 237
Db 542 ACTGAGGAAGCATATCTCTTACCTGGACTAGATGGCACCTGCACAAATTTCTTCAGAAAAT 601
QY 238 ValGlyValLysValIleAspSerIleAsnIleThrLeuGlyAlaGluAspGluLeuLys 257
Db 602 ATTGGTGTTCAGTCTCGACTCTGTGAATATACCTGGGTGGTCCATGATTTCCGA 721
QY 258 HisAlaValGlyLeuValArgProValSerValAlaPheGluValLysGlyPheAsn 277
Db 662 CATGCAGTTGCATTTGTTCTCCAGTCAGTGAGTGTCGATTTGAGTTCATGATTTCCGA 721
QY 278 LeuTyrLysValGlyValTyrSerSerAspThrCysGlyArgAspProMetAspValAsn 297
Db 722 TTTTACAGAAAGAGATTACACAGTGGAACTTTGTACAGCACTCCCATCGGATGGAAC 781
QY 298 HisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIleProTyr 313
Db 782 CATGCTGTTCTTCAGTCGGGTATGGAGTTGGATCTGCTGTCATCATAC 829

RESULT 5
CK295860
LOCUS
DEFINITION
EST758574 Nicotiana benthamiana mixed tissue cDNA library,
normalized, full-length Nicotiana benthamiana cDNA clone NBMCX76 5'
end, mRNA sequence.
CK295860
ACCESSION
CK295860.1 GI:39880659
VERSION
KEYWORDS
SOURCE
ORGANISM
Nicotiana benthamiana
Nicotiana benthamiana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

```

REFERENCE AUTHORS TITLE JOURNAL COMMENT

asterids; lamids; Solanales; Solanaceae; Nicotiana.
1 (bases 1 to 977)
Buell, C.R., Hart, A., Zismann, V., Karamycheva, S.A., Day, B.,
Staskawicz, B., Jin, H. and Baker, B.
Generation of EST sequences from Nicotiana benthamiana
Unpublished (2003)
Other ESTs: EST758575
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
Clones can be requested from TIGR via potato@tigr.org
Seq primer: ATT TAG GTG ACA CTA TAG.

FEATURES source

1..977
/organism="Nicotiana benthamiana"
/mol_type="mRNA"
/db_xref="taxon:4100"
/clone="NBMCX76"
/tissue_type="abiotic and biotic stress-treated leaves,
callus tissue and root tissue"
/lab_host="DH10B-Tona"
/clone_lib="Nicotiana benthamiana mixed tissue cDNA
library, normalized, full-length"
/note="Vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;
supplier: RNA was isolated from Nicotiana benthamiana
tissues that include callus, roots from liquid culture
grown plants, heat-stressed leaves (38 C, 3 hr and 6 hr),
cold-stressed leaves (5 C 3 hr, 6hr), and pathogen
challenged leaves (Pseudomonas syringae pv tomatu 12 hr;
Xanthomonas campestris pv campestris 12 hr, 18hr;
Pseudomonas syringae pv phaseolicola 18hr, and Xanthomonas
campestris pv vesicatoria 18hr). RNA was isolated from
these tissues and pooled in approximately equal molar
amounts."

ORIGIN

Alignment Scores:
Pred. No.: 2,21e-112 Length: 977
Score: 1034.00 Matches: 208
Percent Similarity: 76.51% Conservative: 33
Best Local Similarity: 66.03% Mismatches: 65
Query Match: 55.44% Indels: 9
DB: 14 Gaps: 5
US-10-087-714-2 (1-352) x CK295860 (1-977)
QY 5 LeuLeuPhePheLeuLeuPheLeuValSerAlaLeuSerValAlaLeuAlaGly----- 22
Db 33 CTCCTATTGGCTCTCGTCGCGCGTGGCTTTTCGCGCGCACTTGGCGACCGGT 92
QY 23 ---PheGluGluAspAsnProIleArgSer---ValThrGlnArgProAspSerIleGlu 40
Db 93 ACCTTTCGCGGATGAGAAATCCGATCAGACAAATTTCTTCGCGTTTACATGAGCTGGAG 152
QY 41 ProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAla 60
Db 153 AACGGAATTCCTCAAGTGTGCGCAAGACTGTCATGCTCTCTCTTCGCTCGCTTGTCT 212
QY 61 ArgArgTyrGlyLysSerTyrGlySerGluGluGluLysLysArgPheGlyIlePhe 80
Db 213 CACAGTATGGGAAGAGGTACGACAGCTTGGAGAGATAAAGCAAGGTTCCGAGTATTC 272
QY 81 ValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyrThrLeuGly 100
Db 273 TTGGCAATTTGAAGATGATTCGATCGCATAAACAAAGGACTATCATCAAACTCGGT 332
QY 101 IleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAlaAla 120
Db 333 GTCATGAGTTTACCGACATAAATGAGGACGAGTTCCGAGACAGACAGGTTGGGAGCAGCT 392
QY 121 GlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProVal 140

```

Db      393 CAAAACTGTTTCAGCCACCACAAAGGGCAATCTCAAACCTCACTAAACGTTGCTGCGCGAG 452
QY      141 ThrArgAspTTPArgGluGlnGlylleValSerProValLysAspGlnGlySerCysGly 160
Db      453 ACAGAAAGCTGGAGGAGCTGGGATGTCAGGCCAGTCAAGAACAGGCGCAAGTGCAGGA 512
QY      161 Ser---TTPThrPheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGly 179
Db      513 TCTTGCTGGACATTCAGCACTACTCTGTCACATAGACAGCATATGGCCCAAGCATTTGGG 572
QY      180 Ser-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPhe 197
Db      573 AAGGAATTTCTCTATCTGACACAGCTTGTGGATGTGCTGGAGCTTTTAAATCACTT 632
QY      198 GlyCys----GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGlylle 216
Db      633 GGCTGCAATGGTGGCTCCCATCACAAAGCCTTTGAGTATATTAAATCCATGCTGCTT 692
QY      217 AspThrGluGlnThrTyrProTyrLeuGlyValMetGlylleCysAsnPhelysGlnGlu 236
Db      693 GACACTGAAGACATATCCATACACCGCGCAAGATGGCTTATGTAATTCATCACA 752
QY      237 AsnValGlyValLysValIleAspSerIleAsnIleThrLeuGlyAlaGluAspGlnLeu 256
Db      753 AATGTTGTCATCAAGATCATGATCTCTGCAATATTACCCCTGGGTGCTGAAGATGA 812
QY      257 LysHisAlaValGlyLeuValArgProValSerValAlaPheGluValValLysGlyPhe 276
Db      813 AAATACGCGGTTCATTTGTTAGGCGCTTATGATAGCTTTTCAGGTGATAAAGTTTC 872
QY      277 -AsnLeuTyrIlyLysGlyValTyrSerSerAspThrCysGlyArgAspPrometAspVa 296
Db      873 AAAACAATACAAAGATGGTGTGTACACACACCGCAATGTGGGCATCTCCCATGGATGT 932
QY      296 LasnHisAlaValLeuAlaValGlyTyrGlyValGluAspGly 310
Db      933 AAACCATGCTCTTCTCTGTTGGGTACGGTGTGTTGAATGTT 975

RESULT 6
LOCUS   CF212812
DEFINITION
Vitis vinifera cDNA clone CAST0005_IF_H04 5', mRNA sequence.
ACCESSION
VERSION   CF212812.1 GI:33407185.
KEYWORDS
SOURCE    EST.
ORGANISM  Vitis vinifera
          Vitis vinifera
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          rosids; Vitaceae; Vitis.
REFERENCE
AUTHORS   Goes da Silva,F., Iandolino,A., Lim,H., Baek,J., Jones,K. and
          Cook,D.
TITLE      Expressed sequence tags from Vitis vinifera 'Cabernet sauvignon'
          berries at various developmental stages
JOURNAL   Unpublished (2003)
COMMENT    Contact: Douglas Cook, PhD
          CABES Genome Facility
          UC Davis, Plant Pathology
          One Shields Ave, Davis, CA 95616, USA
          Tel: 530 754 6561
          Fax: 530 754 6617
          Email: drcoc@ucdavis.edu
          Seq primer: ACGGTACGGACATATGCC.
          Location/Qualifiers
            1..1005
              /organism="Vitis vinifera"
              /cult_type="mRNA"
              /multivar="Cabernet Sauvignon"
              /db_xref="taxon:29760"
              /clone="CAST0005_IF_H04"
              /sex="Hermaphrodite"

```

```

/dev_stage="Pre-bloom (10-11 days before bloom)"
/lab_host="DH5alpha"
/clone_lib="Vitis vinifera cv. cabernet sauvignon Stem -
CAST"
/notes="Organ: Stem; Vector: pDNR; Site 1: sf11; Site 2:
sf11; CAST is a cDNA library of Vitis vinifera cv.
'Cabernet Sauvignon' stems. Samples were collected May 13,
2002 from pre-bloom plants (10-11 days before bloom),
pre-veraison. Sampled vines were located at the University
of California, Davis. Experimental vineyard. cDNAs were
made by oligo-dT priming and directionally cloned. 5' and
3' adaptors were used in cloning as follows:
5'-AAGCAGTGTATCAACGAGTGGCCATACGCCCGG-3', and
5'-ATTCTAGAGCGGCGGCGGCGCATG-dt(30)NN-3'. Library was
constructed using the Clontech Creator SMART kit and
size-selected to contain the 0.5-3 kb size fraction."

```

ORIGIN

```

Alignment Scores:
Pred. No.:      5,48e-111      Length:      1005
Score:          1022.50      Matches:      202
Percent Similarity: 74.68%      Conservatives: 31
Best Local Similarity: 64.74%      Mismatches: 61
Query Match:     54.93%      Indels:      18
DB:              14          Gaps:        4

US-10-087-714-2 (1-352) x CF212812 (1-1005)

QY      10 LeuPheLeuValSerAlaLeuSerValAlaLeu----- 20
Db      72 CTCTCGTGCTGCGCGCGCTCTGATCTGCTCTGCCCGCTGCTTCCGGGGAGGTGAC 131
QY      21 -----AlaGlyPheGluGluAspAsnProIleArgSerValThrGlnA-gPro 36
Db      132 CATCATTTCCGATCAAGCTTCGATGAAGAGAACCAATTAGACTGTATCGGACAGCAT 131
QY      37 AspSerIleGluProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPhe 56
Db      192 CCGGACTTGGATGCTCGCTCTCGGCTGATCGGGACAGCGGTCTACTCCTTCT 251
QY      57 AlaArgPheAlaArgArgTyrGlySerTyrGlySerGluGluGluLysLysArg 76
Db      252 GCGAGCTTCGCTTCACAGGTATGGGAAGAGTTACAGACGCGGTGGACGAGATTAAGCTGAGA 311
QY      77 PheGlyIlePheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSer 96
Db      312 TTGAGATTTTCTCGGAGATTTGAACTCATCAGATCCACACACAGAAAGGCTTGCCT 371
QY      97 TyrThrLeuGlylleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArg 116
Db      372 TATACTCTAGCTGTAAATCAGTTTCGCTGATTGGACCTGGGAAGAGTTCCGACACACAGG 431
QY      117 LeuGlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGly 136
Db      432 TTGGGAGCTGCTCAGAACTGCTCTGCCACCTTTGAAGGGCAATCACAAAGCTTAAGTACGTT 491
QY      137 ValLeuProValThrArgAspTyrArgGluGlnGlnIleValSerProValLysAspGln 156
Db      492 ATCTCTCTCGAGACGAAGATGGAGAGAGATGGCATAGTCAGCCCAATCAAGATCAA 551
QY      157 GlySerCysGlySer---TyrThrPheSerThrThrGlyAlaLeuGluAlaAlaTyrThr 175
Db      552 GGTCACTGTGGATCTTGTCTGAGCTTTGAGCACCCTTGAGAGCTCTAGAGGAGCTTACGCT 611
QY      176 GlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAla 193
Db      612 CAGGCATTTGGAGGGGATCTCTCTGTCTGAGCAGCAGCTTTGGAGCTGTGCGGAGCT 671
QY      194 PheAsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLysTyr 212
Db      672 TTCAATAAATTTGGATGCGACCGGGGATGGCATCCCAAGCTTTTGTAGTACATCAATAAC 731
QY      213 AsnGlyGlylleAspThrGluGlnThrTyrProTyrLeuGlyValMetGlylleCysAsn 232

```



```

Db      848  ACCGAGTTTACACGAGCACTGAATGTGGGACACTCCCATCGACGGTAACCATGCTGTT 907
Qy      301  LeuAlaValGlyTyrGlyVal 307
Db      908  CTTGCTGTGGGCTACGGTGT 928

RESULT 8
CK257583      920 bp mRNA linear EST 12-DEC-2003
LOCUS      EST141220 potato callus cDNA library, normalized and full-length
DEFINITION      Solanum tuberosum cDNA clone POC649 5' end, mRNA sequence.
ACCESSION      CK257583
VERSION
KEYWORDS
SOURCE      EST.
ORGANISM      Solanum tuberosum (potato)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum.
REFERENCE      1 (bases 1 to 920)
AUTHORS      Buell, C.R., Hart, A., Zismann, V., Karameycheva, S.A. and Baker, B.
TITLE      Generation of ESTs from potato callus tissue
JOURNAL      Unpublished (2003)
COMMENT      The Institute for Genomic Research
              9712 Medical Center Dr, Rockville, MD 20850, USA
              Email: potato-array@tigr.org
              Clones can be requested from TIGR via potato@tigr.org
              Seq primer: ATT TAG GTG ACA CTA TAG.

FEATURES
    source
        Location/Qualifiers
            1..920
                /organism="Solanum tuberosum"
                /mol_type="mRNA"
                /cultivar="Kennebec"
                /db_xref="taxon:4113"
                /clone="POC649"
                /tissue_type="callus"
                /lab_host="DH10B-Tona"
                /clone_lib="potato callus cDNA library, normalized and
                full-length"
                /note="vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;
                supplier: RNA was isolated from Solanum tuberosum var.
                Kennebec callus tissue grown on solid media."

ORIGIN
Alignment Scores:
Pred. NO.:      6.8e-109      Length:      920
Score:          1004.50      Matches:    204
Percent Similarity: 76.97%      Conservative: 30
Best Local Similarity: 67.11%      Mismatches: 61
Query Match:    53.86%      Indels:     9
DB:             14      Gaps:        6

US-10-087-714-2 (1-352) x CK257583 (1-920)
Qy      9  LeuLeuPheLeuValSerAlaLeu---SerValAlaLeuAlaGly-----PheGlu 24
Db      11  CTATGGTCTTGTGCGCGCTTTTCGCTCTGCTGCTGCTGCGGACCGGACCTTCGCC 70
Qy      25  GluAspAsnProIleArgSerValThrGlnArgProAspSerIleGluProAlaIleLeu 44
Db      71  GATGAGATCCGATCAGCAAGTAGTAGTT---ACCGAAGAGCTGGAGACGGAAATCTTT 127
Qy      45  GlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAlaArgTyrGly 64
Db      128  CAAGTCGTGGCCAGACTCGGAATGCTCTCTCTGCTGCTGCTGCTGCTATCAGGACCGG 187
Qy      65  LysSerTyrGlySerGluGluIleLysArgPheGlyIlePheValGluAsnLeu 84
Db      188  AAAAGGTACGAGCTCGGTGAGGAGATCAAGCAAGAGGTTTCGAGATATTTTGGACAATCTG 247
Qy      85  AlaPheIleArgSerThrAsnArgLysAspLeuSerTyrThrLeuGlyIleAsnGlnPhe 104

```

```

Db      248  AAGATGATCCGATCGCATAAACAGCAAGGACTATCATCAAACTCGGTGTCATGAGTTT 307
Qy      105  AlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAlaAlaGlnAsnCysSer 124
Db      308  ACGACCTTAACATGGGATGAGTTCGCTAGACACAAGATTGGGGGATCTCAAAACTGTCT 367
Qy      125  AlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProValThrArgAspTyr 144
Db      368  GCCACTACAAAGGCAATCTCAAGCTAACTAAGCTGCTTCTGCCAGAGAGAGGACTGG 427
Qy      145  ArgGluGlnGlyIleValSerProValLysAspGlnGlySerCysGlySer---TrpThr 163
Db      428  AGGGAAGATGGTATTGTAGCCAGTGAAGGCACAGGGCAAGTGGCGATCTTGCTGGACA 487
Qy      164  PheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGlySer-----Thr 181
Db      488  TTCAGCACTACTGTGCTAGAGGCGAGCATATGCCCAAGCATTTGGGAAGGGGAATCTCT 547
Qy      182  LeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPheGlyCys---Gly 200
Db      548  CTGTCAGAGCAGCAGCTTGTGGACTGTGCTGGACCTTTTAATACTTTGGCTGCAATGG 607
Qy      201  GlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGlyIleAspThrGluGln 220
Db      608  GGGTGGCATCACAAAGCCTTTGAGTACATTAAATTCATGCTGTGCTTGACACACTGAAGA 667
Qy      221  ThrTyrProTyrLeuGlyValMetGlyIleCysAsnPheLysGlnGluAsnValGlyVal 240
Db      668  GCATATCCATACACCGGCAAGATGGCATATGTAATTCACAAAGCAAAATATTGGTGTC 727
Qy      241  LysValIleAspSerIleAsnIleThrLeuGlyAlaGluAspGluLeuLysHisAlaVal 260
Db      728  AAAGTCATCAGTTCTGTCAATATTACCTGGTGTGCTGAAGATGAATCAATATCGCAGTT 787
Qy      261  GlyLeuValArgProValSerValAlaPheGluValValLysGlyPheAsnLeuTyrLys 280
Db      788  GCATGGTGTAGCCCTGTAGTGTTCCTTTGAGGTGTAAAGGTTTCAACACAGTATAAG 847
Qy      281  LysGlyValTyrSerSerAspThrCysGlyArgAspProMetAspValAsnHisAlaVal 300
Db      848  ACGGAGATTACACCGACACTGAATGTGGCGACACTCCCATCGGACGTAAACCATGCTGTT 907
Qy      301  LeuAlaValGly 304
Db      908  CTTGCTGTGGGC 919

```

RESULT 9

CB982399

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

748 bp mRNA linear EST 01-MAY-2003
 CB982399
 CAB70006_IVaF_A04 Cabernet Sauvignon Berry Post-Veraison - CAB7
 Vitis vinifera cDNA clone CAB70006_IVaF_A04 5', mRNA sequence.

CB982399
 CB982399.1 GI:30305605

EST.
 Vitis vinifera

Vitis vinifera

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; Vitaceae; Vitis.

1 (bases 1 to 748)

Goes da Silva, F., Iandolino, A., Lim, H., Baek, J., Jones, K. and

Cook, D.

Expressed sequence tags from cabernet sauvignon berries at various

developmental stages

Unpublished (2003)

Contact: Douglas Cook, PhD

CAES Genome Facility

UC Davis, Plant Pathology

One Shields Ave, Davis, CA 95616, USA

Tel: 530 754 6561

Fax: 530 754 6617

Email: drccook@ucdavis.edu

Seq primer: ACGTACCGGACATATGCC.

FEATURES		Location/Qualifiers	
source		1..748	
		/organism="Vitis vinifera"	
		/mol_type="mRNA"	
		/cultivar="Cabernet Sauvignon"	
		/db_xref="taxon:29760"	
		/clone="CAB70006 IVaF_A04"	
		/sex="Hermaphrodite"	
		/dev_stage="Post-Veraison, 18-19 brix"	
		/lab_host="DHSalpa"	
		/clone_lib="Cabernet Sauvignon Berry Post-Veraison - CAB7"	
		/note="Organ: Berry; Vector: pDNR; Site 1: SfiI; Site 2: SfiI; CAB7 is a cDNA library of Cabernet Sauvignon Clone 8 berries. Samples were collected post-veraison from field-grown vines during stage III of berry growth at 89 days after full bloom. Berries soluble solid concentration ranged between 18-19 brix. Sampled vines were located at the University of California, Davis, Experimental Vineyard. cDNAs were made by oligo-dT priming and directly cloned. 5' and 3' adaptors were used in cloning as follows: 5'-AAGCAGTGTATCAACGAGTGGCCATTACGCGCGG-3' and 5'-ATTCTAGGCGGCGGCGGCGGATG-3' (30' NN-3'. Library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5-3 kb size fraction."	
ORIGIN			
Alignment Scores:			
Pred. No.:	4.67e-107	Length:	748
Score:	988.00	Matches:	187
Percent Similarity:	81.93%	Conservative:	17
Best Local Similarity:	75.10%	Mismatches:	41
Query Match:	52.98%	Indels:	4
DB:	14	Gaps:	3
US-10-087-714-2 (1-352) x CB982399 (1-748)			
Qy	72	GlutLeuLysArgPheGlyValGluAsnLeuAlaPheIleArgSerThrAsn	91
Db	2	GAGATTAGCTGAGATTCGAGATTTCGAGAGATTGAACTCATCATCCCAAC	61
Qy	92	ArgLysAspLeuSerThrLeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGlu	111
Db	62	AGAAAGGCTTGCCTATACCTAGCTGTTAATCAGTTCGCTGATTCGACCTGGGAAG	121
Qy	112	PheArgThrAsnArgLeuGlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHis	131
Db	122	TTCCGCACACAGGTGGGAGCTGCTCAGAACGCTCTGCACCTTGAAGGGCAATCAC	181
Qy	132	ArgPheValAspGlyValLeuProValThrArgAspTrpArgGluGlnGlyIleValSer	151
Db	182	AAGCTAAGCTGAGTTATCTTCTCAGACGAAAGATTGGAGAGATGGCATATTTCAGC	241
Qy	152	ProValLysAspGlnGlySerCysGlySer---TrpThrPheSerThrThrGlyAlaLeu	170
Db	242	CCAAATCAAGATCAAGGTCACTGTGGATCTTGTCTGGACTTTTCAGCACCACTGGAGCTTA	301
Qy	171	GluAlaAlaThrThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuVal	188
Db	302	GAGGAGCTTACGCTCAGCGATTGGGAGGGGATCTCTCTCTGAGCAGCAGCTGTG	361
Qy	189	AspCysAlaSerAlaPheAsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPhe	207
Db	362	GACTGTGCGGAGCTTTCAATAACTTTGATGTCACGCGGGGATTCGCCATCCCAAGCTTTT	421
Qy	208	GlnThrValLysThrAsnGlyIleAspThrGluGlnThrThrProThrLeuGlyVal	227
Db	422	GAGTACATCAATAACAAAGGTGGTGGCTTGATCTAGGAGGAGCATATCTTACCTGGACTA	481
Qy	228	MetGlyIleCysAsnPheLysGlnGlnAsnValGlyValLysValIleAspSerIleAsn	247
Db	482	GATGGCACTGCAAAATTTCTTCAGAAAAATATTGGTGTTCAGTCTTCGACTCTGTGAAT	541
Qy	248	IleThrLeuGlyAlaGluAspGluLeuLysHisAlaValGlyLeuValArgProValSer	267
Db	56	PheAlaArgPheAlaArgArgThrGlyLysSerThrGlySerGluGluLeuLysLys	75
Db	7	TTCCGCTCTTTCCTATCATCGGCATCGAAAGTACGAGTCCGCTTTCGAGGATCAAGCAA	66

QY 76 ArgPheGlyLeuPheValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeu 95
 Db 67 AGGTTGAGATATTTGGCAATCTCGATGATCCGATCGCATACAGCAAGGATTA 126
 QY 96 SerTyrThrLeuGlyLeuGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsn 115
 Db 127 TCATACAACTCGGTCTCAATGATGTTACCGACCTAACATGGATGAGTTCGTAGACAC 186
 QY 116 ArgLeuGlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAsp 135
 Db 187 AAGTTGGGGCATCTCAAACTCTTCTGACACTACAAAGGGCAATCTCAAGTAACACTAAC 246
 QY 136 GlyValLeuProValThrArgAspTrpArgGluGlnGlyLeuValSerProValLysAsp 155
 Db 247 GTCGTTCTGCGACAGACGAGGACTCGAGGGAAGATGTTATGTTAGCCAGTGAAGGCA 306
 QY 156 GlnGlySerCysGlySer---TrpThrPheSerThrGlyAlaLeuGluAlaLys 174
 Db 307 CAGGCAAGTGGCGATCTTCTGGACATTCAGCACTACTGGTGCTAGGACGACATAT 366
 QY 175 ThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSer 192
 Db 367 GCCCAAGCATTTGGGAAGGGAATCTCTCTGTCAGACGACGAGCTTGTGGACTGTGCTGA 426
 QY 193 AlaPheAsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrValLys 211
 Db 427 GCTTTTAAATAACTTTTGGCTGCAATGGGGGTTGCCATCAAGCCCTTGAATACATTA 486
 QY 212 TyrAsnGlyGlyLeuAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyLeuCys 231
 Db 487 TTCAATGGTGGTCTTGACACTGAAGAAGCATATCCATACACCGCAAGAATGGCATATGT 546
 QY 232 AsnPheLysGlnGluAsnValGlyValLysValIleAspSerIleAsnIleThrLeuGly 251
 Db 547 AAATTCTCAACCAAAATATTTGGTGTCAAAGTCATCAGTCTCTCAATATTACCCCTGGGT 606
 QY 252 AlaGluAspLeuLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGlu 271
 Db 607 GCTGAAGATGAATGAAATACGAGTTCATTTGGTTAGGCTGTAGTGTCTTTTGG 666
 QY 272 ValValLysGlyPheAsnLeuTyrLysLysGlyValTyrSerSerAspThrCysGlyArg 291
 Db 667 GTGGTAAAGGTTTCAACACAGTATAAGACGGAGTTTACACGACGACTGAATGTGGCGAC 726
 QY 292 AspProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyIle 311
 Db 727 ACTCCATGAGCACTAAACCATGCTGCTGTCTGTGGGTACGGTGGTGAATGGCGTT 786
 QY 312 ProTyrTrpLeuIleLysAsnSerTrp 320
 Db 787 CCCTACTGGTCTCAT-AAGAACTCATGG 812
 RESULT 11
 CK294622
 LOCUS CK294622 965 bp mRNA linear EST 15-DEC-2003
 DEFINITION EST757336 Nicotiana benthamiana mixed tissue cDNA library,
 normalized, full-length Nicotiana benthamiana cDNA clone NEMCO76 5',
 end, mRNA sequence.
 ACCESSION CK294622
 VERSION CK294622.1 GI:39878196
 KEYWORDS EST.
 SOURCE Nicotiana benthamiana
 ORGANISM Nicotiana benthamiana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; lamiids; Solanales; Solanaceae; Nicotiana.
 REFERENCE 1 (bases 1 to 965)
 AUTHORS Buell, C.R., Hart, A., Zismann, V., Karamycheva, S.A., Day, B.,
 Skaskawicz, B., Jin, H., and Baker, B.
 TITLE Generation of EST sequences from Nicotiana benthamiana
 JOURNAL Unpublished (2003)
 COMMENT Contact: Robin Buell

The Institute for Genomic Research
 9712 Medical Center Dr. Rockville, MD 20850, USA
 Email: potato-array@tigr.org
 Clones can be requested from TIGR via potato@tigr.org
 Seq primer: AT1 TAG GIG ACA CTA TAG.
 FEATURES
 source
 1. 965
 /organism="Nicotiana benthamiana"
 /mol_type="mRNA"
 /db_xref="taxon:4100"
 /clone="NEMCO76"
 /tissue_type="abiotic and biotic stress-treated leaves,
 callus tissue and root tissue"
 /lab_host="DH10B-Tona"
 /library="Nicotiana benthamiana mixed tissue cDNA
 library, normalized, full-length"
 /note="Vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;
 supplier: RNA was isolated from Nicotiana benthamiana
 tissues that include callus, roots from liquid culture
 grown plants, heat-stressed leaves (38 C, 3 hr and 6 hr),
 cold-stressed leaves (5 C 3 hr, 6hr), and pathogen
 challenged leaves (Pseudomonas syringae pv tomato 12 hr;
 Xanthomonas campestris pv campestris 12 hr, 18hr;
 Pseudomonas syringae pv phaseolicola 18hr, and Xanthomonas
 campestris pv vesicatoria 18hr). RNA was isolated from
 these tissues and pooled in approximately equal molar
 amounts."
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.72e-106 Length: 965
 Score: 981.00 Matches: 195
 Percent Similarity: 75.67% Conservative: 32
 Best Local Similarity: 65.00% Mismatches: 65
 Query Match: 52.60% Indels: 8
 DB: 14 Gaps: 5
 US-10-087-714-2 (1-352) x CK294622 (1-965)
 QY 5 LeuLeuPhePheLeuLeuPheLeuValSerAlaLeuSerValAlaLeuAlaGly----- 22
 Db 64 CTCCTATTGGCTCTCGTCGTCGCGCGGTGGCGCTTTTCGCCCGCCGACCTTGGCGACCGGCT 123
 QY 23 ---PheGluGluAspAsnProIleArgSer---ValThrGlnArgProAspSerIleGlu 40
 Db 124 ACCITTCGCGATGAGATCCGATCAGACAAATTTTCTCAGCGGTTTACATGAGCTGGAG 183
 QY 41 ProAlaIleLeuGlyValLeuGlySerCysArgHisAlaPheHisPheAlaArgPheAla 60
 Db 184 AACGGAATTTCTCAAGTCGTCGCGCAAGACTCGTCATGCTCTCTCTTCGCTCGCTTTTGT 243
 QY 61 ArgArgTyrGlyLysSerTyrGlySerGluGluGluIleLysLysArgPheGlyIlePhe 80
 Db 244 CACAGGTATGGAGAGGTACGACAGACAGTTCGAGAGATAAAGCAAGGTTTCGAGGTATTC 303
 QY 81 ValGluAsnLeuAlaPheIleArgSerThrAsnArgLysAspLeuSerTyrThrLeuGly 100
 Db 304 TTGACAAATTTGAAGATGATTTCGATCGCATACAAAGAAAGGACTATCATCAAACTCGGT 363
 QY 101 IleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGlyAlaAla 120
 Db 364 GTCATGAGTTTACCGACATAACATGGAGAGAGTTCGCGAGACAGAGTTGGGAGCAGCT 423
 QY 121 GlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProVal 140
 Db 424 CAAAACTGTTTCAGCCACCAAGGCAATCTCAAACTCACTAACGTTGTCTCTCGCGAG 483
 QY 141 ThrArgAspTrpArgGluGlnGlyIleValSerProValLysAspGlnGlySerCysGly 160
 Db 484 ACGAAGACTGGAGGGAAGCTGGGATTTGTACGCCAGTCAAGACCAAGGCAAGTGGCGGA 543
 QY 161 Ser---TrpThrPheSerThrThrGlyAlaLeuGluAlaLysTyrThrGlnLeuThrGly 179

Db 544 TCTTCTGGACATTCAGCAGCTACTGGTGCACCTAGAACGACGATATGGCCAAAGCATTTTGGG 603

Qy 180 Ser-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAspPhe 197

Db 604 AAGGAAATTTCTCTATCTGAACAGCAGCTTGGAGCTGCTGGAGCTTTTAATAACTTT 663

Qy 198 GlyCys---GlyGlyLeuProSerGlnAlaPheGluThrValLysTyrAsnGlyGlyLe 216

Db 664 GCCTCAATGGTGGCTCCCATCAACAGCTTTGAGTATATAAATCAATGGTGGTCTT 723

Qy 217 AspThrGluGlnThrTyrProTyrLeuGlnValMetGlyLeuCysAsnPheLysGluGlu 236

Db 724 GACACTGAAGAGCATATCCATACACCGCAAGATGGCTTATTAATTTCTATCAGAA 783

Qy 237 AsnValGlyValLysValLysSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 256

Db 784 AATGTGTGTGTCAAGATCATGATCTCTCAATATTACCTGGGTGTGTAAGATGAAC 843

Qy 257 LysHisAlaValGlyLeuValArgProValSerValAlaPheGluValValLysGlyPhe 276

Db 844 AAATACGGGTGTCATTGGTTAGCCCTTTAGTATAGCTTTTGGAGTGATAAAGGTTTC 903

Qy 277 AsnLeuTyrLysLysGlyValTyrSerSerAspThrCysGlyArgAspProMetAspVal 296

Db 904 AAACAATACAAGATGGTGTATTACACGACGACCAATGTGGCCATCTCCCATGGATGA 963

RESULT 12

LOCUS BQ281381

DEFINITION BQ281381.1 731 bp mRNA linear EST 13-MAY-2002

WHE3020_F06_L12S Wheat unstressed seedling shoot normalized cDNA

library Triticum aestivum cDNA clone WHE3020_F06_L12, mRNA

sequence.

ACCESSION BQ281381

VERSION BQ281381.1 GI:20548869

KEYWORDS EST

SOURCE Triticum aestivum (bread wheat)

ORGANISM Triticum aestivum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

AUTHORS Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

TITLE Poideae; Triticeae; Triticum.

JOURNAL 1 (bases 1 to 731)

COMMENT Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D.,

US Department of Agriculture, Agriculture Research Service, Pacific

West Area, Western Regional Research Center

800 Buchanan Street, Albany, CA 94710, USA

Tel: 5105595773

Fax: 5105595818

Email: candersn@pw.usda.gov

Sequences have been trimmed to remove vector sequence and low

quality sequence with phred score less than 20

Seq primer: SK primer.

Location/Qualifiers

1..731

/organism="Triticum aestivum"

/mol_type="mRNA"

/cultiivar="Chinese Spring"

/db_xref="taxon:4585"

/clone="WHE3020_F06_L12"

/tissue_type="Etiolated shoot"

/dev_stage="Five day old seedling"

/lab_host="E. coli DH10B"

/clone_lib="Wheat unstressed seedling shoot normalized

cDNA library"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid

pBluescript SK, Site 1: EcoRI; Site 2: XhoI; Seeds were

surface-sterilized, Germinated and grown aseptically in

the dark at room temperature on filter paper with water,

nystatin and cefotaxime in covered crystallization dishes. Roots were harvested. The tissue, total RNA, and poly(A) RNA were prepared, a cDNA library was made in the 10 Close lab (Choi, Close, Fenton) at the University of California, Riverside. The cDNA clones were in vivo excised to give pBluescript phagemids before normalization was carried out. The mass excision of phagemid library and normalization were done in HT Nguyen lab by D. Zhang at Texas Tech University. Normalization protocol used was that of Soares'. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

ORIGIN

Alignment Scores: 1,41e-105 Length: 731

Pred. No.: 975.50 Matches: 182

Score: 86.36% Conservative: 27

Percent Similarity: 75.21% Mismatches: 28

Best Local Similarity: 52.31% Indels: 5

Query Match: 13 Gaps: 4

DB:

US-10-087-714-2 (1-352) x BQ281381 (1-731)

Qy 78 GlyLeuPheValGluAsnLeuAlaPheLeuArgSerThrAsnArgLysAspLeuSerTyr 97

Db 4 GGGATCTTCTCCGAGAGCCTCGAGCTGCTCCACCAACCGAGGGGCGCTCCCTAC 63

Qy 98 ThrLeuGlyLeuAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeu 117

Db 64 CGCTTCGGCATCAACCGTTTCGCGACATGAGCTGGAGGAGTTCCAGCGAGCGGCTC 123

Qy 118 GlyAlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAsp---Gly 136

Db 124 GCGCGCGGCGAGACTGCTCGCGACGCTCGCGGCAACCAACCGATGCGCGACGCGCC 183

Qy 137 ValLeuProValThrArgAspTrpArgGluGlnGlyLeuValSerProValLysAspGln 156

Db 184 GCTCTCCGCGACCAAGACTGGAGGAGGATCGTGAGCCCTGAGGCCCTGTAAGACCAAG 243

Qy 157 GlySerCysGlySer---TrpThrPheSerThrThrGlyAlaLeuGluAlaAlaTyrThr 175

Db 244 GGTCACTGTGTCTCTGTGGACCTTCAGCACCACTGTTCTCTTCTTGGGCAGCATATACT 303

Qy 176 GlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAla 193

Db 304 CAGGCACCTGGAAACCTGCTCTCTTTCTGAGCAGCAGCTGGTGTGTGTGCTACTGCA 363

Qy 194 PheAsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluThrValLysTyr 212

Db 364 TACAATAATTTCCGATGCAGTGGAGGCTTACCATCTCAGGCTTTGAGTACATCAATAC 423

Qy 213 AsnGlyGlyLeuAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyLeuGlyCysAsn 232

Db 424 AATGTGCGCTTGACACAGAGAGCTTACCTTACAGCGGTGTCAATGGCATCTGTCTAT 483

Qy 233 PheLysGlnGluAsnValGlyValLysValLysSerLeuAsnLeuThrLeuGlyAla 252

Db 484 TACAAGCTGAAAGAGTTGGAGTCAAGTTTGGACTCCGTTACATCACTCCCTGGTGCT 543

Qy 253 GluAspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPheGluVal 272

Db 544 GAGGATGAGCTGAAGAATGCTGTGTGGAGCTTGTTCGTCAGTTAGTTGTCCTTTCAGGTG 603

Qy 273 ValLysGlyPheAsnLeuTyrLysGlyValTyrSerSerAspThrCysGlyArgAsp 292

Db 604 ATCAACGTTTCAGGATGTACAGAGTGGAGTTTACACAGTACCATTGGGAACTTCT 663

Qy 293 ProMetAspValAsnHisAlaValLeuAlaValGlyTyrGlyValGluAspGlyLeuPro 312

Db 664 CCAATGATGTGAACCAACGCTGTCTCGCCCTGTGTATGTCGTAAGAAATGGGTTCCC 723

Qy 313 TyrTrp 314

```

Db      724 TACTGG 729
|||||
RESULT 13
CB892658
LOCUS   768 bp mRNA linear EST 24-APR-2003
DEFINITION HOGA Medicago truncatula cDNA clone HOGA-19K6, mRNA
sequence.
ACCESSION CB892658
VERSION   CB892658.1 GI:30099826
KEYWORDS EST.
SOURCE   Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 768)
AUTHORS Hahn,M.G., Ojanen-Reuhs,T., Samac,D., Town,C.D., Van Aken,S.,
Utterback,T., Cho,J. and Fraser,C.M.
TITLE ESTs from roots of Medicago truncatula treated with
oligogalacturonides of DP 6-20
JOURNAL Unpublished (2001)
COMMENT Contact: Michael G. Hahn
Complex Carbohydrate Research Center
University of Georgia
220 Riverbend Road, Athens, GA 30602-4712, USA
Tel: 706-542-4457
Fax: 706-542-4412
Email: hahn@ccrc.uga.edu
TIGR sequence name: WTMCV637K
More information is available at: www.medicago.org
Seq primer: SKnoc (CTA GAA CTA gtg gat CC).
FEATURES
source
1..768
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="A17"
/db_xref="taxon:3880"
/clone="HOGA-19K6"
/tissue_type="3 day old seedling roots"
/dev_stage="24 hours after treatment in the dark at 26 C
with 0.5 mg/ml oligogalacturonides (DP 6-20) in the
presence of 100 ug/ml Gentamicin"
/lab_host="XLOLR"
/clone_lib="HOGA"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the Unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in SOLR cells."
ORIGIN
Alignment Scores:
Pred. No.: 6.9e-105 Length: 768
Score: 970.00 Matches: 181
Percent Similarity: 81.25% Conservative: 27
Best Local Similarity: 70.70% Mismatches: 44
Query Match: 52.01% Indels: 4
DB: 14 Gaps: 3
US-10-087-714-2 (1-352) x CB892658 (1-768)
QY 65 LysSerTyrGlySerGluGluIleLysLysArgPheGlyIlePheValGluAsnLeu 84
Db 1 AAAAGATACGATCCGCTGATGAATGAAGCGTAGATTTAAGATCTTCTGAAAATCTT 60
QY 85 AlaPheIleArgSerThrAsnArgLysAspLeuSerTyrThrLeuGlyIleAsnGlnPhe 104
Db 61 CAACCTTATCAATCTACTAATAAGAAACCCGCTGGTTATCTCTCGGTGTTAATCATTTT 120

```

```

QY 105 AlaAspLeuThrTyrGluGluPheArgThrAsnArgLeuGlyAlaAlaGlnAsnCysSer 124
Db 121 GCTGATTGGACTTGGGAGGAGTTTCAAGAGTCATAGACTCGGTGCTGCTCAAAATTTCT 180
QY 125 AlaThrAlaHisGlyAsnHisArgPheValAspGlyValLeuProValThrArgAspTyr 144
Db 181 GCTGCTCTCAAGGGAACCATAGGATTACCGATGTTTCTTCCCGCTGAGAAAGACTGG 240
QY 145 ArgGluGlnGlyIleValSerProValLysAspGlnGlySerCysGlySer---TyrThr 163
Db 241 AGAAAGAAGAGTATAGTCAGTGAAGTTAAGATCAAGGCCACTGCGGATCATGTGGACA 300
QY 164 PheSerThrThrGlyAlaLeuGluAlaAlaTyrThrGlnLeuThrGlySer-----Thr 181
Db 301 TTCAGCAACACTGGAGCTTTGGAGTCAGCTTACGCACAGCGCTTCGGAAGAATATCTCT 360
QY 182 LeuSerGluGlnGlnLeuValAspCysAlaSerAlaPheAsnAsnPheGlyCys---Gly 200
Db 361 CTTTCTGAGCAGCAGCTAGTAGACTGTGCTGGTGTCTTACAATACTTTGGCTGCAATGGT 420
QY 201 GlyLeuProSerGlnAlaPheGluTyrValLysTyrAsnGlyGlyIleAspThrGluGln 220
Db 421 GGGTTGCCATCCCAAGCCTTTGAATACATTAAATACAAATGCTGGCTTGAGACAGAGAA 480
QY 221 ThrTyrProTyrLeuGlyValMetGlyIleCysAsnPheLysGlnGluAsnValGlyVal 240
Db 481 GCATATCCCTACACTGGCAAAATGGTCTCTGCAAAATTACATCTGAAAAGCTTGGCGTT 540
QY 241 LysValIleAspSerIleAsnIleThrLeuGlyAlaGluAspGluLeuLysHisAlaVal 260
Db 541 CAAGTCTCTGGCTCTGTCAATATCACCTGGCTGCTGAGGATGAATGAACATGCAAGTT 600
QY 261 GlyLeuValArgProValSerValAlaPheGluValValLysGlyPheAsnLeuTyrLys 280
Db 601 GCTTTTGTCTGGCCCGTGTAGTGGCAATTCAGGTGGTGTGATGACTTCAGTTATACAAG 660
QY 281 LysGlyValTyrSerSerAspThrCysGlyArgAspProMetAspValAsnHisAlaVal 300
Db 661 AAAGAGTTTACACTAGTACAACTTGTGGCAGCACACCCATGATGATGTAATCATCTGCT 720
QY 301 LeuAlaValGlyTyrGlyValGluAspGlyIleProTyrTrpLeuIle 316
Db 721 CTTGCTGTGGTATGAATGAAGATGGTGATGATGATGATGATGATGATGATGATGAT 768
RESULT 14
LOCUS BE195255 794 bp mRNA linear EST 22-OCT-2001
DEFINITION HVSMEH0088L13f Hordeum vulgare 5-45 DAP spike EST library
HVCN00099(5 to 45 DAP) Hordeum vulgare subsp. vulgare cDNA clone
HVSMEH0088L13f, mRNA sequence.
ACCESSION BE195255
VERSION BE195255.3 GI:16321152
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 794)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D.,
Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W.,
Fenton,R.D., Close,S.J., Oates,R. and Main,D.
Development of a genetically and physically anchored EST resource
for barley genomics: Morex 5-45 DAP spike cDNA library
Unpublished (2001)
JOURNAL On Jun 26, 2000 this sequence version replaced gi:13188047.
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu

```

Total hg bases = 385
 Seq primer: AATTAACTCTCACTAAAGG
 High quality sequence stop: 489.
 Location/Qualifiers
 1. 794
 /organism="Hordeum vulgare subsp. vulgare"
 /mol_type="mRNA"
 /cultivar="Morex"
 /sub_species="vulgare"
 /db_xref="taxon:112509"
 /clone="HVSMEH008813f"
 /tissue_type="5-45 DAP Spike"
 /lab_host="SOUR"
 /clone_lib="Hordeum vulgare 5-45 DAP spike EST library
 HVCNA0009 (5 to 45 DAP)"
 /note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;
 Plants were grown in the greenhouse at the University of
 California, Riverside (Fenton, SJ Close, TJ Close). Whole
 spikes with awns trimmed were collected at 5, 10, 15, 20,
 30 and 45 DAP (Fenton). Total RNA was prepared from each
 pool, equal quantities of all six RNA pools were combined,
 poly(A) RNA was purified from the mixture, one primary
 unamplified cDNA library was made, and 1 million pfu were
 in vivo excised to give pBluescript SK(-) cDNA phagemids
 (Choi) in the TJ Close lab at the University of California,
 Riverside. Phagemids were plated and picked at the Clemson
 University Genomics Institute (CUGI) (Begum, Palmer,
 Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
 sequencing and sequence analysis were performed at CUGI
 (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
 The sequence has been trimmed to remove vector sequence
 and contains a minimum of 100 bases of phred value 20 or
 above. For more details on library preparation and
 sequence analysis see
<http://www.genome.clemson.edu/projects/barley>. To order
 this clone see <http://www.genome.clemson.edu/orders> Also
 see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
 Genetically and physically anchored EST resources for
 barley genomics. Barley Genetics Newsletter 31:29-30.
 (<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)"

FEATURES

source

ORIGIN

Alignment Scores:
 Pred. No.: 1,448-104 Length: 794
 Score: 967.50 Matches: 183
 Percent Similarity: 81.71% Conservative: 27
 Best Local Similarity: 71.21% Mismatches: 40
 Query Match: 51.88% Indels: 7
 DB: 10 Gaps: 4

US-10-087-714-2 (1-352) x BE195255 (1-794)

Qy 99 LeuGlyIleAsnGlnPheAlaAspLeuThrTrpGluGluPheArgThrAsnArgLeuGly 118
 Db 6 CTCGGCATCACCGTCTCGACATGAGCTGGAGAGATTCCAGGCGCCGCTCTCGGC 65
 Qy 119 AlaAlaGlnAsnCysSerAlaThrAlaHisGlyAsnHisArgPheValAsp---GlyVal 137
 Db 66 GCGGCGCAGACCTGCTCGGCGACGCTCGCGGCAACACCTCATCGGAGCGCCGCG 125
 Qy 138 LeuProValThrArgAspTrpArgGluGlnGlyIleValSerProValIleAspGlnGly 157
 Db 126 CTCCTGGAGACCAAGACATGGAGGAGGATGGGATCGTCAGCCCGCTCAAAAACCGAGCC 185
 Qy 158 SerCysGlySer---TrpThrPheSerThrGlyAlaLeuGluAlaAlaThrGln 176
 Db 186 CATTGGGCTCTGCTGGCGGTTGACACTCTGCGGCATCTGAGCAGCATATCTACTCAG 245
 Qy 177 LeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAlaSerAlaPhe 194
 Db 246 GCCACTGGAAGAACAATCTCTCTTCTGAGCAACAGCTGGTGTGACTGTGCTGGTGGTTC 305
 Qy 195 AsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrrValIleValAsn 213

Db 306 AATACTTTGGATGCAACGAGGCGCTTCCATCCAGCGCTTTGAGTACATCAATCAAT 365
 Qy 214 GlyGlyIleAspThrGluGlnThrTrpTyrrLeuGlyValMetGlyIleCysAsnPhe 233
 Db 366 GGGGGGATCGACACTGAGGAGTCTACCTCTTACAGGGTGTCAATGGCGTCTGCCATTAC 425
 Qy 234 LysGlnGluAsnValGlyValIleValIleAspSerIleAsnIleThrLeuGlyAlaGlu 253
 Db 426 AAGCTGAATAATGCTCGGTTTCAAGTTTGGACTCAGTAACATCACACTGAATGCCGAG 485
 Qy 254 AspGluLeuIleHisAlaValGlyLeuValArgProValSerValAlaPheGluValVal 273
 Db 486 GATGAATGTAAGAACGCGCTCGGTTTGGTTCGCCAGTCAGGCTGGCTTTCAGTGATT 545
 Qy 274 LysGlyPheAsnLeuTyrrIleValIleValIleValIleValIleValIleValIle 293
 Db 546 GACGGTTTACGAGTATACAGACCGAGTTTACCAACAGCAGCATTTGTGGCATTACCCCT 605
 Qy 294 MetAspValAsnHisAlaValLeuAlaValGlyTyrrGlyValGluAspGlyIlePro-Ty 313
 Db 606 GGTGACGTGAACCAACCGCTTCTTGGCAGTCGGCTATGGTGTGAGAACGGGCGTCCGGTA 665
 Qy 313 rTrpLeuIleIleAsnSerTrp-GlyThrAsnTrpGlyAspAsnGlyTyrrPheIleValMetG 333
 Db 666 TTGGCTTATCAAGAACTCGTGGGGGCGCGACTGGGGTGACAAATGGGTACTTCAAGATGA 725
 Qy 333 luLeuGlyIleValAsnMetCysGlyValAlaThrCysAlaSerTyrrPro 348
 Db 726 AATGGGAGAACATGTGCTATTGGTAAGTGGCTTCTACCCCT 772

RESULT 15
 BE131652
 LOCUS
 DEFINITION
 L48-1651T3 Ice plant Lambda Uni-Zap XR expression library, 48 hours
 NaCl treatment Mesembryanthemum crystallinum cDNA clone L48-1651,
 mRNA sequence.
 ACCESSION
 BE131652
 VERSION
 BE131652.1 GI:8579015
 KEYWORDS
 EST.
 SOURCE
 Mesembryanthemum crystallinum (common iceplant)
 Mesembryanthemum crystallinum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Caryophyllales; Alzooceae; Mesembryanthemum.
 REFERENCE
 1 (bases 1 to 720)
 Cushman, J.C.
 TITLE
 An expressed sequence tag database for the common ice plant,
 Mesembryanthemum crystallinum
 JOURNAL
 Unpublished (1997)
 COMMENT
 Contact: Cushman JC
 Department of Biochemistry
 University of Nevada
 MS200, Reno, NV 89557-0014, USA
 Tel: 775-784-1918
 Fax: 775-784-1650
 Email: jcushman@unr.edu
 PCR Primers
 FORWARD: T7
 BACKWARD: T3
 Plate: L48-17
 Seq primer: T3
 High quality sequence stop: 350
 POLYA=No.

FEATURES
 source
 1. 720
 /organism="Mesembryanthemum crystallinum"
 /mol_type="mRNA"
 /db_xref="taxon:3544"
 /clone="L48-1651"
 /tissue_type="Leaf, 48 h 0.4M NaCl"
 /dev_stage="Six week old"
 /clone_lib="Ice plant Lambda Uni-Zap XR expression

library, 48 hours NaCl treatment"
/note=Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
EcoRI; Site_2: XhoI"

ORIGIN

Alignment Scores:
Pred. No.: 2,24e-103 Length: 720
Score: 957.00 Matches: 179
Percent Similarity: 83.26% Conservative: 20
Best Local Similarity: 74.90% Mismatches: 36
Query Match: 51.31% Indels: 4
DB: 10 Gaps: 3

US-10-087-714-2 (1-352) x BE131652 (1-720)

```

QY 115 AsnArgLeuGlyAlaAlaGlnAsnCysSerAlaAlaThrAlaHisGlyAsnHisArgPheVal 134
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 3 CACAGATTGGAGCTGCCCAAACTGCTGTCACACTCAAGGGCAGTCACAGCTGTCT 62

QY 135 AspGlyValLeuProValThrArgAspTTPArgGluGlnGlyLeuValSerProValLys 154
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 63 CATGCAGACCTTCTCGAAACAAAAGATTGGAGAGCTGAGGGCATAGTCAAGCCGGTCAAG 122

QY 155 AspGlnGlySerCysGlySer---TTPThrPheSerThrThrGlyAlaLeuGluAlaAla 173
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 123 AATCAACACCACTGTGGATCTCTGGACATTCAGTACAACTGGAGCTTTGGAGGAGCT 182

QY 174 TyrThrGlnLeuThrGlySer-----ThrLeuSerGluGlnGlnLeuValAspCysAla 191
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 183 TATGGCGAGGCAATTTGGGAAGAACATCTCTTCTGAGCAACAACTTGTGATTGTGCT 242

QY 192 SerAlaPheAsnAsnPheGlyCys---GlyGlyLeuProSerGlnAlaPheGluTyrVal 210
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 243 GGTGCATTTGATTAATCATGTTGCAATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 302

QY 211 LysTyrAsnGlyGlyIleAspThrGluGlnThrTyrProTyrLeuGlyValMetGlyIle 230
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 303 AAGTACAATGGAGGACTTGACACAGAGAGGCTTATCTTACACAGCAAGGATGGTGA 362

QY 231 CysAsnPheLysGlnGluAsnValGlyValLysValIleAspSerIleAsnIleThrLeu 250
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 363 TGTAAATTTCTCAGCCCCAAAATGTTGGTGTTCAGTCTCTGGACTCCGTTAACATCACCC 422

QY 251 GlyValaGluAspGluLeuLysHisAlaValGlyLeuValArgProValSerValAlaPhe 270
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 423 GAGCTGAAGAGGAACTTAAGGATGCTGTGGCATTTGGTCCGTCCTGTGAGTGTGGATT 482

QY 271 GluValValLysGlyPheAsnLeuTyrLysGlyValTyrSerSerAspThrCysGly 290
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 483 GAGGTTGTTAATGGATTCCGGTTTACACGAAAGAGATTACACCAAGCACTTGTGGC 542

QY 291 ArgAspProMetAspValAsnHisAlaValAlaValGlyTyrGlyValGluAspGly 310
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 543 AGCAGCCCCATGATGTGAACCATGCTGTTCTAGCTGTGGATGTGGAGTGGATGGT 602

QY 311 IleProTyrTrpLeuIleLysAsnSerTrpGlyThrAsnTrpGlyAspAsnGlyTyrPhe 330
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 603 ATCCATATCTGGCTTATCAAGAACTCGTGGGAGAGATTGGGGTGACATGGATGACTTC 662

QY 331 LysMetGluLeuGlyLysAsnMetCysGlyValAlaThrCysAlaSerTyrProIle 349
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 663 AAGATGGAGATGGGAAAAAACAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 719

```

Search completed: March 30, 2004, 00:10:17
Job time : 2379 secs